Genetics of resistance to *Fusarium* crown rot in barley and pan-genome sequence anchors of this species

by

Shang Gao

M.S. in Genetics and Plant Breeding

B.S. in Biology

Submitted in fulfilment of the requirements for

the degree of Doctor of Philosophy

University of Tasmania

January 2020

Declarations of Originality

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Signed:

Date: 21 Jan 2020

Authority of Access

The publishers of the four papers, comprising Chapter 3 to 6 of this thesis, hold the copyright for that content and access to the material should be sought from the respective journals. The remaining nonpublished content of the thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Signed:

Date: 21 Jan 2020

Statement of co-authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

Candidate	Shang Gao	Tasmanian Institute of Agriculture, UTAS
Author 1	Chunji Liu	Agriculture and Food, CSIRO
Author 2	Meixue Zhou	Tasmanian Institute of Agriculture, UTAS
Author 3	You-Gan Wang	School of Mathematical Science, QUT
Author 4	Zhi Zheng	Agriculture and Food, CSIRO
Author 5	Jinran Wu	School of Mathematical Science, QUT
Author 6	Jiri Stiller	Agriculture and Food, CSIRO
Author 7	Jonathan Powell	Agriculture and Food, CSIRO
Author 8	Ahsan Habib	Agriculture and Food, CSIRO
Author 9	Yunfeng Jiang	Agriculture and Food, CSIRO
Author 10	Miao Liu	Agriculture and Food, CSIRO
Author 11	Haiyan Hu	Agriculture and Food, CSIRO
Author 12	Haoran Shi	Triticeae Research Institute, SICAU
Author 13	Jian Ma	Triticeae Research Institute, SICAU
Author 14	Yaxi Liu	Triticeae Research Institute, SICAU
Author 15	Yuming Wei	Triticeae Research Institute, SICAU
Author 16	You-Liang Zheng	Triticeae Research Institute, SICAU

Abbreviations:
UTAS: University of Tasmania
CSIRO: The Commonwealth Scientific and Industrial Research Organisation
QUT: Queensland University of Technology
SICAU: Sichuan Agricultural University

Contribution of work by co-authors for each paper:

PAPER 1: Located in Chapter 3

Gao, S., Zheng, Z., Hu, H., Shi, H., Ma, J., Liu, Y., Wei, Y., Zheng, Y., Zhou, M. and Liu, C., 2019. A novel QTL conferring Fusarium crown rot resistance located on chromosome arm 6HL in barley. Frontiers in Plant Science, 10:1206

Author contributions:
Conceived and designed experiment: Author 1, Author 2, Author 15, Author 16
Performed the experiments: Candidate, Author 4, Author 11
Analysed the data: Candidate, Author 12, Author 13
Wrote the manuscript: Candidate, Author 1, Author 4, Author 16

PAPER 2: Located in Chapter 4

Gao, S., Zheng, Z., Powell, J., Habib, A., Stiller, J., Zhou, M. and Liu, C., 2019. Validation and delineation of a locus conferring Fusarium crown rot resistance on 1HL in barley by analysing transcriptomes from multiple pairs of near isogenic lines. BMC Genomics 20: 650.

Author contributions:
Conceived and designed experiment: Author 1, Author 2
Performed the experiments: Candidate, Author 4, Author 8
Analysed the data: Candidate, Author 6, Author 7
Wrote the manuscript: Candidate

PAPER 3: Located in Chapter 5

Gao, S., Zheng, Z., Hu, H., Jiang, Y., Liu, M., Stiller, J., Zhou, M., Liu, C., 2020. Delineating a locus conferring Fusarium crown rot resistance on chromosome arm 1HL in barley by developing and analysing a large population derived from near isogenic lines. The Crop Journal. https://doi.org/10.1016/j.cj.2020.03.008

Author contributions:
Conceived and designed experiment: Author 1, Author 2
Developed materials: Candidate, Author 9, Author 10
Performed the experiments: Candidate, Author 4, Author 11,
Analysed the data: Candidate
Wrote the manuscript: Candidate, Author 1

PAPER 4: Located in Chapter 6

Gao, S., Wu, J., Stiller, J., Zheng, Z., Zhou, M., Wang, Y., and Liu, C., 2020. Identifying barley pan-genome sequence anchors using genetic mapping and machine learning. Theoretical and Applied Genetics, https://doi.org/10.1007/s00122-020-03615-y

Author contributions:
Conceived and designed experiment: Candidate, Author 1, Author 3
Cleaned, analysed and visualized the data: Candidate
Trained the machine learning model: Candidate, Author 5
Developed the scripts: Candidate, Author 6
Conducted the association analysis: Candidate, Author 4
Wrote the manuscript: Candidate

We, the undersigned, endorse the above stated contribution of work undertaken for each of the published peer-reviewed manuscripts contributing to this thesis:
Signed:

Shang Gao Candidate Tasmanian Institute of Agriculture University of Tasmania Date: 21 January 2020	Professor Meixue Zhou Primary Supervisor Tasmanian Institute of Agriculture University of Tasmania Date: 21 January 2020	Professor Michael Rose Director Tasmanian Institute of Agriculture University of Tasmania Date: 11 February 2020

Acknowledgements

It has been a long journey to complete my PhD candidature, and I sincerely thank all those who have given me valuable support along the way. Particularly, for the harmonious and pleasant working environment provided by the staff, visitors and students working with me in St Lucia, CSIRO. For his friendship and his hard work in our collaborations – Dr Zhi Zheng. For his kind, smart and skilful help in bioinformatics – Dr Jiri Stiller. For their ever-present love, the warmest care and constant encouragement – my farther Guixin Gao, my mother Baoqin Zhao and my wife Lily Jin.

I would like to thank for the financial support from the University of Tasmania (UTAS) and China Scholarship Council (CSC). I would also like to express my sincere gratitude to Professor Meixue Zhou from UTAS who give me the great chance and unreserved support to carry out my PhD project under his supervision.

Lastly, I would like to say the biggest ‘thank you’ to Dr. Chunji Liu from CSIRO for his excellent co-supervision, his care, his wise guidance, his patience and integrity and the generous sharing of his expertise. I believe, so far, it is the most fortunate experience for me to study in Chunji's team and his critical thinking, humble courtesy and extraordinary perseverance will keep guiding me in the rest of my life.

Table of contents

List of Figures

List of Tables

Abstract

Fusarium crown rot (FCR), caused primarily by *Fusarium pseudograminearum*, is a devastating disease for cereal production in semi-arid regions worldwide. Apart from severe yield loss, this disease can also lead to accumulation of mycotoxins in foods or feeds, which is harmful to health of human and livestock. It has long been recognized that breeding and growing resistant varieties is an important and indispensable component in FCR management.

Three large-effect quantitative trait loci (QTL) conferring FCR resistance had been identified in barley when I first started my PhD research. They were located on the chromosome arms 1HL, 3HL and 4HL, respectively. Pyramiding these QTL was proved to be effective in further enhancing the resistance to FCR. I started my PhD research by characterizing a new source of resistance. It (AWCS799) is a landrace identified from a systematic screening of more than 1,000 genotypes. Genetic control of its resistance was investigated by generating and analysing two populations of recombinant inbred lines with AWCS799 as the common parent. One of the populations was used for QTL detection and the other for validation. A novel QTL, located on the long arm of chromosome 6H (designated as *Qcrs.caf-6H*), was consistently detected in each of the four tests conducted against the mapping population. The QTL explained up to 28.3% of the phenotypic variance and its effect was confirmed in the validation population. Significant interaction between this resistance locus and either plant height or heading date was not detected in the populations used, further facilitating its manipulation in breeding programs.

The interactions between FCR severity and other characteristics indicate that QTL detected through mapping can only be treated as putative. Benefiting from the uniform genetic backgrounds for untargeted genomic region, near isogenic lines (NILs) could be used to validate the effectiveness of QTL for various characteristics. Validating an existing locus on chromosome arm 1HL thus become another part of my PhD research program. This QTL is named as *Qcrs.cpi-1H*. Five pairs of NILs targeting this locus were generated. Analysing the NILs found that the resistant allele at *Qcrs.cpi-1H* significantly reduced FCR severity. Transcriptomic analysis was then conducted against three of the NIL pairs, which placed the 1HL locus in an interval spanning about 11 Mbp. A total of 56 expressed genes bearing single nucleotide polymorphisms (SNPs) were detected in this interval. Five of them contain non-synonymous SNPs. NILs developed in this study and the transcriptomic sequences obtained

from them would be valuable for identifying genes and generating diagnostic markers targeting this locus.

As the limited resolution caused by the heterogeneous genetic backgrounds in mapping populations, markers obtained from QTL mapping is usually not tightly linked with a given locus. To develop markers that are able to reliably trace 1HL locus in breeding programs, we developed and assessed a fine mapping population consisting of 1,180 recombinant inbred lines derived from one of the above NIL pairs. Using this population, we delineated *Qcrs.cpi-1H* into an interval of 0.4 cM covering a physical length of about 487 kb. Six markers co-segregating with this locus were generated. Three of the five genes with non-synonymous variations identified from the multiple pairs of NILs were further confirmed to be located within the interval. In addition to generate extra markers for breeding programs, the refined location of *Qcrs.cpi-1H* should also facilitate the cloning of the causal gene(s) underlying this locus.

It has become clear in recent years that many genes in a given species cannot be found in any given genotype, thus a comprehensive pan-genome can be highly valuable for genetic research and breeding. Clearly, obtaining a comprehensive pan-genome requires deep-sequencing large numbers of genotypes which is still not practical for species like barley which has a huge and highly repetitive genome. However, large quantity of genotype-by-sequencing (GBS) data has been made available for barley. We thus attempted to identifying barley pan-genome sequence anchors using these data based on an approach combining genetic mapping and machine learning. Based on the GBS sequences from 11,166 domesticated and 1,140 wild barley genotypes, we identified 1.844 million reliable tags. Of them, 532,253 were identified as presence/absence variation (PAV) tags. Based on those present in the genome of the hulless barley genotype Zangqing320, positions for 83.6% of the Morex-absent tags from the domesticated genotypes and 88.6% from the wild barley genotypes should be correct. Association analyses against flowering time, plant height and kernel size showed that the relative importance of the PAV and Non-PAV tags varied for different traits.

This project provided a series of genetic resources, including a novel locus and the detailed genetic profile of *Qcrs.cpi-1H*, for breeding FCR resistant barley varieties. Moreover, the high-resolution physical map based on pan-genome sequences should not only facilitate the construction of a comprehensive barley pan-genome, but also assist various genetic studies including identification of structural variation, genetic mapping and breeding in barley.

Chapter 1 General introduction

Fusarium crown rot (FCR) is a chronic disease in cereal crops mainly caused by *Fusarium pseudograminearum*. It is prevalent throughout the world in arid and semi-arid crop areas (Akinsanmi et al. 2004; Chakraborty et al. 2006). The disease in Australia (Murray and Brennan 2010; Murray and Brennan 2009) and the USA (Smiley et al. 2005b) was known to cause significant yield loss. In recent years, FCR has also become a major cereal production issue in China (Li et al. 2016; Xu et al. 2017; Zhou et al. 2014).

FCR pathogens are carried in crop residues and can live in the field environments for more than two seasons (Chakraborty et al. 2006; Smiley et al. 2005b) making it hard to control the disease only using farming practices such as crop rotation (Burgess 2014). As known several decades ago, growing resistant varieties is an indispensable part of efficiently disease management (Purss 1966; Wildermuth and Purss 1971). The method of detecting and integrating large-effect QTL into elite cultivars has been successfully used in the breeding of varieties with high level of FCR resistance in wheat (Zheng et al. 2017) and barley (Chen et al. 2015). Up to the beginning of this project, three major QTL conferring FCR resistance have been identified in barley which locate on chromosome arms 1HL (Chen et al. 2013b), 3HL (Li et al. 2009) and 4HL (Chen et al. 2013a), respectively.

Gene pyramiding multiple QLT has proven effective in significantly enhancing FCR resistance in barley (Chen et al. 2015). Obviously, only three loci conferring FCR resistance is far from enough for barley breeding program and expanding the genetic resources that impart FCR resistance is urgently needed. Towards identifying novel QTL with strong FCR resistance, a QTL mapping research against an unstudied resistant landrace, selected from a systematic screen of FCR-inoculation (Liu et al. 2012b), formed the first part of this thesis.

Clear interactions have been well documented in barley between FCR severity and agronomic traits including flowering time (Chen et al. 2013a; Liu et al. 2012a) and plant height (Li et al. 2009; Liu et al. 2010). These interactions suggest that the FCR QTL detected from QTL mapping study can only be treated as putative. Thus, the effectiveness of a QTL detected from segregating populations needs to be validated. Near-isogenic lines (NILs) have been used widely in validating QTL for various characteristics (Chen et al. 2014; Pumphrey et al. 2007b). As the two isolines for NILs pair share a similar genetic background and the difference between them is largely caused by the distinct targeted locus. The FCR resistance conferred by 4HL

locus has been validated by multiple pairs of NIL s in a previous work (Habib et al. 2016). In this project, we developed a series of NIL pairs for the 1HL locus (1H NIL pairs) to validate its effectiveness.

RNA-sequencing has been a routine tool mainly used to decipher the differentially expressed patterns of transcriptomes in the past decade (Mortazavi et al. 2008; Stark et al. 2019; Wang et al. 2009). The technique is now commonly used to identify genetic variants for diverse purposes as well (Blencowe et al. 2009; Cavanagh et al. 2013). It has been repetitively certified that RNA-sequencing multiple NILs for a certain QTL is helpful for narrowing down the candidates of causal genes underlying the locus and developing genetic markers which could be used for further delineating the targeted locus (Habib et al. 2018; Jiang et al. 2019; Ma et al. 2014). In this project, three 1H NIL pairs were RNA-sequenced for detecting the transcriptomic responses of hosts to FCR infection and developing genetic makers for further fine mapping the 1HL locus.

It has been well known that markers generated from QTL mapping studies cannot always accurately trace a locus in marker-assisted selection, due to the limited resolution provided by such studies (Paterson et al. 1988). One of the main obstacles is the segregations of undesirable traits in a typical QTL mapping population. As described previously, plant height and growth rate could severely affect FCR severity in barley (Bai and Liu 2015; Chen et al. 2013a; Chen et al. 2014; Li et al. 2009; Liu et al. 2010). Variances of such traits in a mapping population would distort phenotypic evaluation of FCR severity. This drawback could be solved by using a population derived from NILs which segregates mainly for the targeted locus. Additionally, through expanding the population size, the NIL-derived population could generate adequate number of recombinants within the candidate region which is another basis for developing markers tightly linked with the given locus. In this thesis, the 1HL FCR QTL was delineated into a refined interval using a NIL-derived population and a series of co-segregated markers with this locus were developed.

With the progress in plant genome sequencing, there is a growing clear scope that a single genome reference cannot represent all genetic elements in a biological clade. Such a phenomenon can be described by the concept of pan-genome, initially proposed in bacterial genomes (Tettelin et al. 2005), which consist of core genome (genetic elements shared by all genotypes) and dispensable genome (genetic elements present in partial genotypes or even only one genotype). Ma et al. (2019) reported that the dispensable genes account for 38% in barley

(*Hordeum vulgare L.*) and the proportion would increase along with the increase in the numbers of genotypes sequenced.

So far, the genetic mapping studies on FCR were all based only on the international reference genome, i.e. Morex (Mascher et al. 2017). It is understandable that the genes within the candidate genomic region of every QTL may not be fully represented by the Morex reference. A comprehensive pan-genome comprised pseudomolecule-scale individuals are the most ideal tools for identifying dispensable genetic elements within candidate region of a QTL. However, that is still not practical to obtain in-depth sequences for numerous barleys which has a large (about 5.1 Gb) (International Barley Genome Sequencing Consortium 2012) and highly repetitive genome (Wicker et al. 2017). In this project, a high-resolution pan-genomic physical map was constructed as a pilot work of building a comprehensive barley pan-genome based reduced sequencing data of over 12,000 genotypes.

In summary, the objectives of this PhD project were to:

1. Characterize a novel source of FCR resistance by QTL mapping;
2. Validate novel QTL detected in the mapping exercise by developing and assessing multiple pairs of NILs as well as transcriptomes from them;
3. Delineate a validated locus by developing and analysing a large population derived from the NILs;
4. Identify barley pan-genome sequence anchors using genetic mapping and machine learning.

Chapter 2 Literature review

2.1 Barley

2.1.1 Origin

Barley (*Hordeum vulgare ssp. vulgare*), domesticated from its wild ancestor *Hordeum vulgare ssp. spontaneum*, is the fourth most important cereal crop worldwide after rice, wheat and maize (FAOSTAT 2018: http://www.fao.org/faostat). Plant remains of barley non-shattering rachises were found in an archaeological site in Israel which dated from at least 10,000 years (Weiss and Zohary 2011). Those barley remains are likely the earliest known of plants morphologically modified by human selection (Allard 1999). It has been mainly recognized that barley was initially domesticated and cultivated at the Israel-Jordan area and then spread and diversified in different eco-geographical region worldwide (Badr et al. 2000; Zeng et al. 2018; Zohary et al. 2012). Nowadays, numerous commercial barley varieties have now been developed to meet diverse requirements on growing season, end uses and morphological differences (e.g. hulled or hull-less, six or two-row seeds) by stringent breeding and selection.

2.1.2 Genome sequencing

Cultivated barley is diploid (2n = 14). It has a huge and complex genome with an estimated size of ~5.1 Gb (Ariyadasa et al. 2014). Until some two decades ago, it had not been feasible to completely sequence and assemble such a huge genome using Sanger sequencing technology due to the prohibitive cost. Aiming at establishing a genomic reference of barley, the barley research community established the International Barley Genome Sequencing Consortium (IBSC) in 2006. As the first step, a BAC-based whole-genome physical map was established due to its immediate value as a fundamental resource for map-based cloning of causal genes underlying different traits (Schulte et al. 2009). Immediately afterwards, a new scope for sequencing the huge barley genome was sparked by second- and next-generation sequencing technologies (Shendure et al. 2017; Shendure and Ji 2008; Shendure et al. 2004) which dramatically reduced overall costs and the time needed for sequencing. These new techniques were adopted by IBSC and they generated a pseudomolecule scaled reference sequence quickly (Mascher et al. 2017). Benefiting from the advances in computational assembly technology, the quality of this barley reference has been continuously improved (Monat et al. 2019b).

Moreover, the high-quality genome assemblies of a few additional genotypes, including cultivated and wild genotypes, have been constructed by various research groups (Dai et al. 2018; Liu et al. 2016b; Zeng et al. 2015). The increased quantity and quality of genome sequences, the improvement of analytical technologies, and the introduction of more research platforms and bioinformatics tools have largely thrived the research and breeding in barley.

2.1.3 World barley production and end uses

In the past several decades, the developments in breeding methods and in farming practices have contributed to a steady increase in barley yield. Nowadays, barley is widely cultivated across both the northern and southern temperate regions (Fig. 2.1a). The annual world barley production is now around 145 million metric tons (Fig. 2.1b).

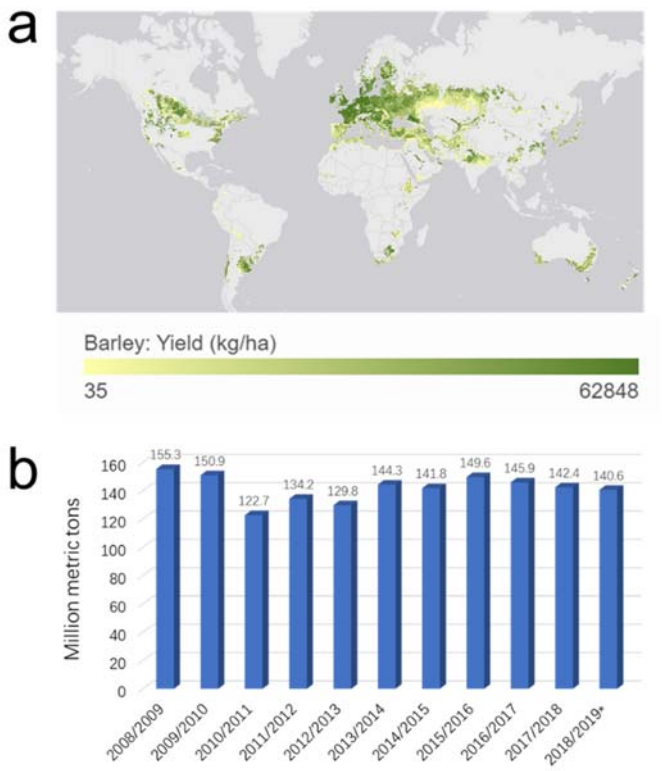


Fig. 2.1 Global barley distribution and yield. **a** Worldwide distribution of barley yield (kg/ha) in 2010. Source: https://gardian.bigdata.cgiar.org/exploration.php#!/ **b** World barley production from 2008/2009 to 2018/2019. The production for 2018/2019 was estimated as of February 2019. Years are from July to June. Source: https://www.statista.com/statistics/271973/world-barley-production-since-2008/

Since 1960, barley has been used most commonly as animal feed, comprising 61 to 77% of the total use of barley (Fig. 2.2). During the same period, the proportion of malting use was increasing from 9% to 22% due to the high economic value and increasing demand for beer production. While barley was originally domesticated for human food and remained as an important source of food in many areas, only around 5% of barley-end use is food consumed in recent years (FAOSTAT 2017).

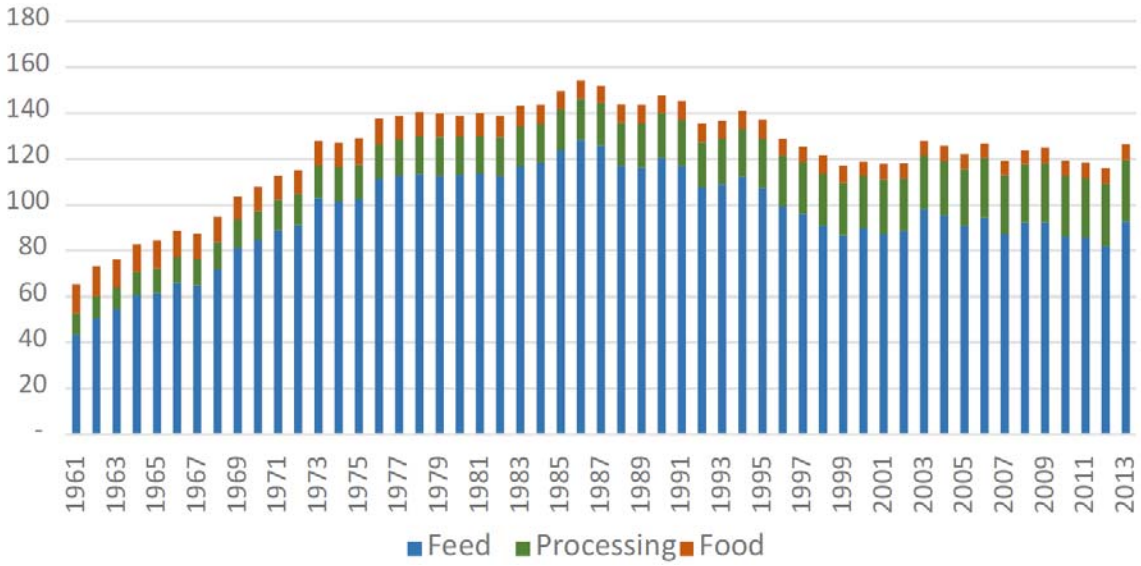


Fig. 2.2 End uses of barley grain in million tonne. The major end uses are for livestock feed (blue), processing or malting (green) and human food (red). Data from FAOSTAT (2017). Cited from "The barley Genome" chapter I, page 4.

2.2 *Fusarium* crown rot

2.2.1 Occurrence areas and damage

Fusarium crown rot (FCR) is a chronic and devastating disease in most semi-arid cereal-producing regions of the world. The occurrence of FCR has been reported in Australia (Murray and Brennan 2010; Murray and Brennan 2009), the Pacific Northwest of the USA (Clear et al. 2006; Smiley and Patterson 1996); Italy, Egypt and Syria (Burgess et al. 2001); at a low frequency in the Mediterranean region and Asia (Bentley et al. ; Tunali et al. 2008). In recent

years, FCR has also become a major issue for cereal production in China (Li et al. 2016; Xu et al. 2017; Zhou et al. 2014).

Smiley et al. (2005) reported that FCR could lead to 35% yield decrease on winter wheat and 13% on barley in the Pacific North-west of the US. In Australia, FCR is endemic over large geographical regions covering the main wheat and barley cropping region. FCR caused about $97 million AUD yield loss in wheat and barley per year (Murray and Brennan 2010; Murray and Brennan 2009). In addition to the loss of yield, FCR infected plants could produce mycotoxins in all tissue including spikes in glasshouse assays (Mudge et al. 2006). The presence of these mycotoxins in food and feed could damage the health of human and living stocks.

2.2.2 FCR pathogen

The pathogen leading to FCR was initially believed to be *Fusarium graminearum Schwabe* (Purss 1966). Based on their morphological divergences and different infection abilities the causal agents were then separated into two groups, group 1 and 2 (Burgess et al. 1975). Group 1 strains were able to cause wheat crown rot, while group 2 strains could give rise to maize stalk rot and wheat head blight. In addition, the two groups could be further distinguished by forming perithecia (group 1) or not (group 2) (Francis and Burgess 1977). Aoki and O'Donnell (1999) reported that the β-tubulin gene sequences of group 1 and 2 could be phylogenetically differentiated. Morphological differences, including colony growth rates, heterothallic production of perithecia, were also identified between these two groups. Based on these evidence, group 1 and 2 were formally named as *Fusarium pseudograminearum* and *Fusarium graminearum*, respectively. Scott and Chakraborty (2006) also demonstrated that *Fusarium pseudograminearum* is a single phylogenetic species and it has also been reported as the globally predominant FCR pathogens in wheat cropping regions (Chakraborty et al. 2006). During infection, *Fusarium pseudograminearum* can produce trichothecene mycotoxin deoxynivalenol (DON) which is a toxin that can inhibit host defense responses to accelerate this process on wheat (Mudge et al. 2006). At the end of a growing season, asexual hyphae and conidia of *Fusarium pseudograminearum* primarily carried by stubble debris turn to be the principle inoculum for the following colonization (Burgess et al. 2001; Paulitz et al. 2002; Summerell et al. 1989).

2.2.3 Symptoms of FCR

The symptoms of FCR show necrotic lesion or browning of leaf sheaths and stem at the beginning of the infection. With the development of this disease, the visible symptoms can also spread into tissues including roots, sub-crown internode, scutellum and coleoptile. In some cases, extreme crown rot infection can cause the seedlings to collapse completely (Purss 1966). In the field, the most obvious indications of FCR on the adult plant can often be observed at the stage of maturity. Once water stress happens during grain filling, the infected plants could encounter early maturation leading to 'white heads' with shrivelled or even no grains (Smiley 2009) (Fig. 2.3a). Another classic symptom on adult plants is honey brown discoloration appears at the tiller base and surrounding lower leaf sheaths (Fig. 2.3b). Under moist conditions, the FCR infected plant can also experience a pink discoloration at the crowns due to the reproduction of *sporodochia* (Fig. 2.3c).

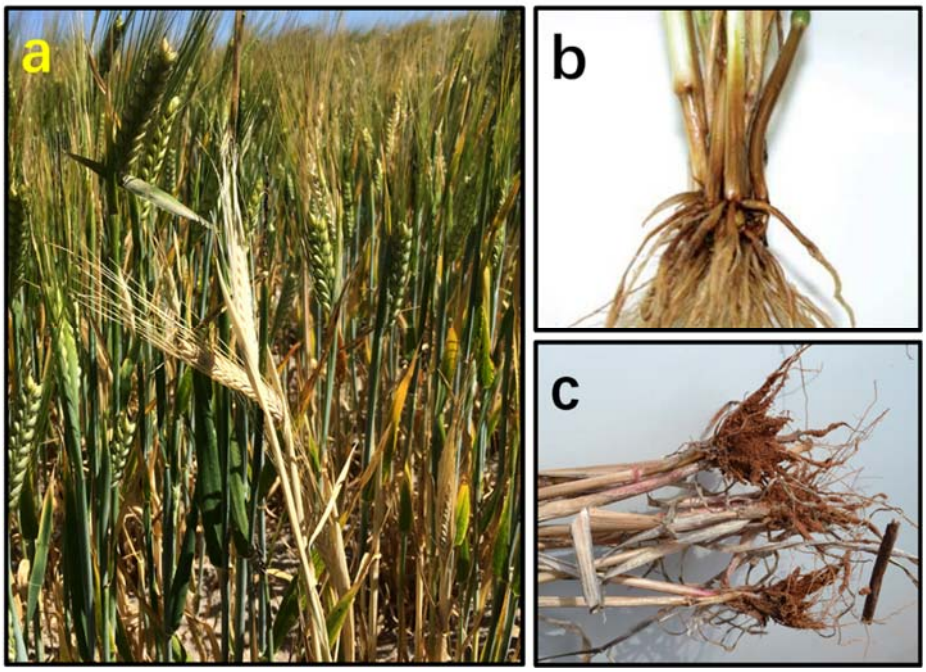


Fig. 2.3 *Fusarium* crown rot (FCR) symptoms caused by *Fusarium pseudograminearum* in barley. **a** FCR caused white heads; **b** A classic symptom of FCR is the honey-brown discoloration of stem bases; **c** Under high moisture, crown rot can also be expressed by a pink discoloration in the crowns of infected cereal plants. Cited from "PestFax" Issue 25, 12 October 2018. (https://agric.wa.gov.au/n/7389)

2.2.4 Factors affecting FCR development

It has widely known that the development of FCR could be influenced by environmental conditions and farming practices. Here, several of the major factors were briefly introduced, including soil moisture, temperature, crop nutrition and planting time and stubble residue.

2.2.4.1 Temperature

It has been documented that temperature can affect the growth and development of *F. pseudograminearum* leading to exacerbation of necrosis lesions on crown rot at both seedlings and adult plants stages. Smiley (2009) discovered that high soil temperature was beneficial for the colonization and infection of *F. pseudograminearum* on wheat due to the strong positive correlation between mean soil temperature and FCR severity. However, Singh et al. (2009) found that the pathogen did not grow when the environmental temperature was over 35 °C with a best suitable range from 10 to 30 °C. This is in basically consistent with the finding made by Backhouse and Burgess (2002) who claimed that the maximum temperatures were just higher than 31 °C in *F. pseudograminearum* detected regions.

2.2.4.2 Soil moisture

The importance of soil moisture for infection and development of FCR has been well recorded (Swan et al. 2000). Based on the observation of a couple of wheat growing seasons, McKnight and Hart (1966) found that the incidence of FCR was higher when precipitation was lower than average. Two studies were conducted using wax partitioned soil columns to investigate the influence of soil-side and seedling-side water potential on FCR infection. Consistent results were obtained which indicated the relatively lower seedling water potential (Beddis and Burgess 1992) and higher soil water potential (Liddell and Burgess 1988) could favor the colonization of FCR pathogen. In the field, the occurrences of whitehead caused by FCR were more often when the soil was dryer at anthesis (Wildermuth et al. 1997).

2.2.4.3 Crop nutrition

Nitrogen concentration in soil has been repeatedly reported as an important factor for FCR infection. Both the usage of fertilizers and crop residues in the field could raise the soil nitrogen levels (Kirkegaard et al. 2004). High soil nitrogen would directly increase the incidence and

severity of FCR (Felton et al. 1998; Verrell et al. 2017). The increasing FCR intensity could be explained by the exacerbation of moisture stress caused by the excessive vegetative growth under high nitrogen level (Burgess et al. 2001). In addition, the soil acidity would increase due to the application of nitrogenous fertilizers and this might be another reason for the aggravation of FCR incidence (Smiley and Patterson 1996).

Zinc is another micronutrient factor for the incidence of FCR on cereal plant. The disease is found to be more severe in the zinc deficit environment (Sparrow and Graham 1988). Grewal et al. (1996) reported that the wheat with higher efficiency of zinc utilization is more resistance to FCR pathogens. Moreover, zinc-deficient could make wheat more susceptible to root rot disease caused by *Rhizoctonia* (Thongbai et al. 1993).

2.2.4.4 Farming practice

Planting time can significantly influence the occurrence of FCR. Purss (1971) reported a higher frequency of FCR incidence in earlier sown plants. A similar result was obtained by Klein et al. (1989) who conduct a field experiment sowing cereal at two different time points in the regions where FCR was epidemic in the previous year. They found that the plants sowed in July suffered less FCR incidence and browning lesions than those sowed in May. However, a recent survey in Australia indicated that late sowing could prolong the crop growth under high temperature and water deficient and finally result in increasing FCR severity and more loss of yield (Simpfendorfer et al. 2012).

As described above, pathogen-infected stubble is the principal inoculum for following FCR occurrence. The causal fungus can survive in residues as mycelium and conidia for two years (Burgess and Griffin 1968). Therefore, the physical contact of newly sowed plants with infected stubble was considered as the main reason for the prevalence of FCR (Burgess et al. 2001; Wearing and Burgess 1977). The incidence of FCR was further exacerbated by the environmentally sustainable management (such as minimum tillage) in Australia (Wildermuth et al. 1997) and more recently in China (Xu et al. 2017).

2.2.5 Agronomic management of FCR

2.2.5.1 Stubble burning

As FCR is a stubble-borne *Fusarium* disease with a broad host range of cereal and grasses, removing the infected plant remains from the field could reduce FCR severity on the following crops. This is confirmed by the studies reported a higher frequency of FCR incidence in stubble retained fields than stubble removed fields (Burgess et al. 1996; Dodman and Wildermuth 1989). Compare to stubble retention, stubble burning could reduce 31.5% occurrence of FCR on average, but the soil nitrogen was decreased by the fires and finally led to low protein content in wheat grains (Summerell et al. 1989). Furthermore, stubble burning could also decrease soil water, damage the beneficial soil microbiome and lead to air pollution (Simpfendorfer et al. 2012). Therefore, the farming system in Australia has given up stubble burning and adopted stubble retention.

2.2.5.2 Crop rotation

Crop rotation is still a pillar for managing FCR (Burgess and Bryden 2012). Planting non-host crops before and after wheat or barley growing season could significantly decrease the available inoculum left in the field (Burgess et al. 2001). Kirkegaard (2004) reported that crop rotation could reduce the incidence of FCR by 3.4 - 41.3% by starving the pathogen left in the field. Sorghum (*Sorghum bicolor*), dry-land cotton (*Gossypium hirsutum*), mungbean (*Vigna radiate*) and canola (*Brassica napus*) are suitable non-host crops for wheat and barley production (Burgess et al. 2001; Simpfendorfer et al. 2012). In terms of the drawbacks, rotation need at least two years to remove infectious hyphae (Hogg et al. 2010) and that may cause economic loss (Liu and Ogbonnaya 2015).

2.3 Breeding FCR resistant varieties

It has been recognized that agronomic managements of FCR incidence may conflict with the economical profits and are not always practical. In contrast, improving the crop genetic-resistance to FCR could overcome the drawbacks of the agronomic practices and lessen FCR damage in cereals as well.

2.3.1 Assessment of FCR resistance

All genetic resistance to FCR can only partially resist the pathogen infestation and they are generally measured according to the severity of FCR symptom and/or pathogen biomass. Given that not only aforementioned environmental factors but also agronomic traits, including plant height (Liu et al. 2010) and growth rate (Chen et al. 2013a), can influence FCR severity, developing a reliable and reproducible method for phenotypic evaluation is the first step of breeding resistant varieties. Through taking a durum wheat as susceptible control and evaluating the FCR severities of a series of rank-known wheat accessions, several different methods have been developed to using glasshouse trial to reliably reflect the FCR performance in the field (Li et al. 2008; Mitter et al. 2006; Wallwork et al. 2004; Yang et al. 2010). Most of these methods take the measurement of browning lesion on the stem base as a proxy for FCR severity to reflect the FCR resistance of accession. Of them, the most serious infections are created by the seedling dip method originally developed by Li et al. (2008) and this method has been employed to detect four QTL in wheat (Ma et al. 2010; Zheng et al. 2014) and three QTL in barley (Chen et al. 2013a; Chen et al. 2013b; Li et al. 2009).

2.3.2 QTL mapping

2.3.2.1 Definition and procedures of QTL mapping

Like many agronomic traits such as plant height and yield, FCR resistance is controlled by multiple genes and present continuously various phenotypic values. Due to the polygenic essence, the resistance to FCR is referred as quantitative resistance. The genomic region containing the genetic element underlying a quantitative characteristic is called quantitative trait loci (QTL) (Miles and Wayne 2008). QTL mapping is developed for detecting the QTL controlling a target trait based on the correlation analysis between phenotypic values of a population and a linkage map of DNA markers constructed from the population.

In general, this method could be conducted as the following procedures: 1) generate suitable populations from the cross between parental lines show significant difference on the trait of interest; 2) genotype the population with feasible genetic markers and calculate a linkage map based on the recombination rate between markers; 3) measure the phenotypic data of all

progenies from the population, and 4) identify the QTL underlying the target trait by statistically analysing the correlation between phenotype and genotype data.

2.3.2.2 QTL conferring FCR resistance in barley

To date, only three QTL conferring FCR resistance have been reported in barley (Liu and Ogbonnaya 2015). The resistance sources from which the QTL were identified were selected from a large-scale screening of genotypes representing diverse geographical origin (Liu et al. 2012b). The QTL on the long arm of 4H was identified from a wild barley (*Hordeum spontaneum*) collected from Iran (Chen et al. 2013a). The other two QTL, located on chromosome arm 1HL and 3HL, were derived from a landrace from Japan (Chen et al. 2013b). Pyramiding three FCR QTL into one genetic background could decrease browning lesions and produce stronger resistance than the effect of a single locus or two different loci (Chen et al. 2015).

2.3.2.3 Limitations of QTL mapping

Apart from the straight-forward concept and numerous successful practices, there are still some factors, including population size, heterozygous genetic background of population and gene-environment interaction, limit the effectiveness of QTL mapping (Asins 2002). Due to the noise generated by these factors, QTL mapping based on segregating DH or RIL populations can only provide limited resolution (Paterson et al. 1988) and the molecular markers developed from QTL mapping cannot always successfully trace the target QTL (Ma et al. 2012). While increasing population size can help to detect low heritability trait to some extant (Beavis 1998), the segregations of unwanted traits in a given population is still a main obstacle for obtaining closely linked marker with a certain locus. This is because the variances of the non-targeted traits can decrease the accuracy of phenotypic assessment of targeted trait (Liu and Ogbonnaya 2015). Therefore, mendelizing the focused QTL, i.e. building a population only segregate at the targeted locus, is an essential step to obtain accurate phenotype data for fine mapping a QTL or even detecting the causal gene underlying the locus. Such a process can be accomplished by the population derived from introgression lines (Paterson et al. 1990) or near-isogenic line (Kølster et al. 1986; Mackill and Bonman 1992).

2.3.3 Near-isogenic lines and RNA-seq

2.3.3.1 Near-Isogenic Lines

Near-Isogenic Lines (NILs) targeting on a given QTL can be generated from either introgression lines (ILs) or heterogeneous inbred families (HIF). Apart from the targeted genomic region from a donor genotype, all other parts of an IL genome are the same with a reference genetic background which is usually taken as a control. ILs can be developed based on backcrossing and selection which begin with phenotype and/or genotype chosen for a donor parental line (Fig. 2.4a). Due to this strategy of development, ILs are also called backcross inbred lines (BILs).

For a large effect QTL of interest, selection can be conducted on the basis on the phenotype of the targeted trait. It is more efficient to integrate marker-assisted selection (MAS) which identified desirable lines based on the genotypic profile of backcross lines (Ribaut and Hoisington 1998). Also, if the dominance of targeted QTL is strong enough, ILs might be directly obtained from backcross lines without selfing generations. Additionally, NILs can also derive from HIF consist of inbred lines which are continuously selfed to F5 or higher generations and still heterozygous at the targeted locus (Fig. 2.4b). With the help of MAS, it is easy to identify the putative lines with homozygous genetic background but heterozygous genomic region of interest (Tuinstra et al. 1997). These putative NILs need to be further confirmed by phenotypic evaluations.

Since NIL pairs are primarily segregated at a specific QTL, the quantitative trait controlled by multiple-locus can be transferred to a Mendelian trait and the phenotypic effect attribute to the given locus can be accurately examined without the noise from other heterozygous genetic elements (Pumphrey et al. 2007a). Recently, combining RNA-sequencing and multiple NILs has been used a new method to identify the candidate genes and possible regulation mechanisms underlying QTL for dormancy in wheat (Barrero et al. 2015), *Fusarium* head blight (FHB) in barley (Huang et al. 2016), FCR in both wheat (Ma et al. 2014) and barley (Habib et al. 2018).

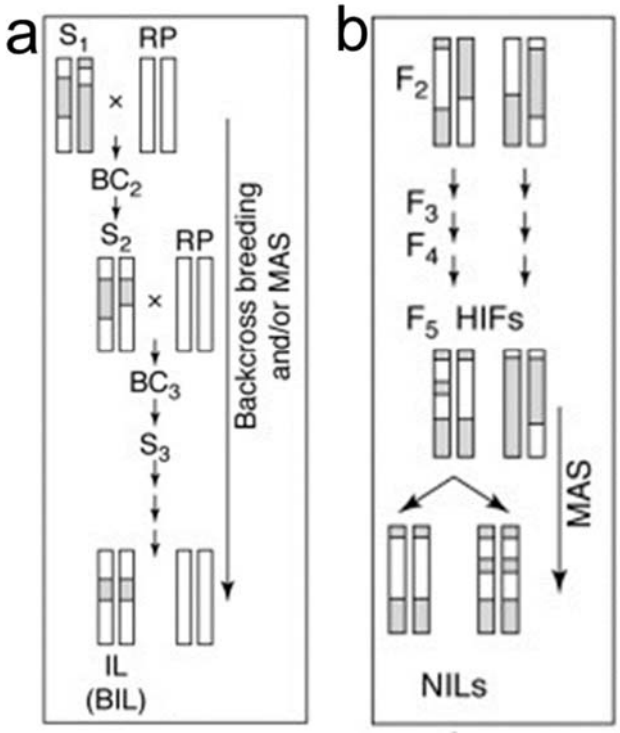


Fig. 2.4 The production of near-isogenic lines (NILs). NILs differing in the genomic region around a targeted QTL can be produce from either **a:** introgression lines (ILs) or **b:** heterogeneous inbred families (HIF). Abbreviations: RP, recurrent parent; S, selected plant; BIL, backcross inbred line; MAS, marker assisted selection. (cited from Alonso-Blanco and Koornneef, 2000 Trends in Plant Science)

2.3.3.2 RNA-sequencing (RNA-seq)

About a decade ago, RNA sequencing (RNA-seq) was first developed to discover genes in Maize (*Zea mays* L.) genome (Emrich et al. 2007) and build the transcriptomic profile of *Arabidopsis thaliana* (Lister et al. 2008). While RNA-seq has been a routine tool in solving problems in various aspects of genomics till date, its major usage is still the analysis of differential expression of genes (DEG) between different samples. The basic steps for DEG analysis, which are still the same with the earliest publications, including library construction, high-throughput sequencing and computational analysis. Library construction successively consists of RNA extraction, mRNA enrichment/ribosomal RNA depletion, cDNA synthesis and adapter ligation. Sequencing step is predominantly based on Illumina short-read platform with 10 to 30 million reads per sample. Computational analysis is aiming to quantify the significant expressional change of transcripts or genes during different treatments or

developmental stages. This step includes aligning the short-reads to a reference, assembling transcripts, counting reads number per transcript, statistically normalize the reads number and analyse the significant expressional changes of transcripts between samples. While the DEG analysis is the main usage of RNA-seq, this technique is also broadly applied to expand our understanding in mRNA splicing, non-coding RNA induced regulation, enhancer RNAs and so on (Stark et al. 2019). For genetic analysis, RNA-seq is also an efficient method to identifying genetic variants, including single nucleotide polymorphism (SNP) and small insertion/deletion (InDel), within transcript or gene region between different genotypes (Cavanagh et al. 2013).

2.3.3.3 Transcriptomic response to FCR infection

Compare with the massive and mounting number of transcriptomic researches on plant disease, only a few studies focused on host transcriptional response to the *Fusarium pseudograminearum*- (*Fp*-) induced crown rot disease in cereals. Taking wheat as a model organism, the defensive responses to biotic stress, anti-microbial pathway and anti-oxidative pathway, were detected in the *Fp*-infected stem base (Desmond et al. 2005). Desmond et al. (2008) reported that both methyl jasmonate and benzothiadiazole, an analogue of salicylic acid, might able to trigger defence to crown rot, as they could activate many *Fp*-induced genes during FCR infection. In addition, deoxynivalenol (DON), a vomitoxin generated during FCR infection, could activate genes coding glucosyltransferases which might be involved in DON detoxification. Such a finding was also reported by Powell et al. (2017a) who presented that uridine-diphosphate glycosyltransferases (UGTs) were possibly involved in detoxification of DON which was important for FCR virulence. A strategy combining RNA-seq and multiple pairs of NILs has been applied on a wheat FCR QTL located at chromosome arm 3BL (Ma et al. 2014) and a barley FCR QTL on chromosome arm 4HL (Habib et al. 2018). These two studies have successfully reduced the candidate genes in tractable numbers and helped further marker development for fine mapping the corresponding FCR QTL.

2.3.4 Fine mapping

As the initial step towards improving breeding efficiency, QTL mapping is unable to generate markers stably tracing a QTL due to the limited resolution (see section 2.3.2.3 for details). Routinely, fine mapping, which is aiming to delineate the causal gene(s) controlling a given trait into a relatively small interval, is necessary for further developing closely linked markers

or identifying the causal gene(s) of a locus. The key factor for fine mapping a locus is building a population which could provide higher mapping resolution than QTL mapping. Such a population is often derived from bi-parental NIL pairs (NIL-derived population) whose genetic backgrounds are highly similar to each other except the heterozygous targeted QTL region (Paterson et al. 1990).

The NIL-derived population allows high-resolution mapping with two reasons. First, phenotypic evaluation could directly detect the variance caused by targeted QTL without background influence (Pumphrey et al. 2007a). Second, informative recombination events, the basis of densifying the genotype profile, could be specifically created within the focused region. Using NIL-derived population consists of 774 lines, Zheng et al. (2015) narrowed a large-effect FCR QTL on the wheat chromosome arm 3BL into a 0.7 cM or 1.5 Mbp. Jiang et al. (2019) delineated a major FCR QTL on the barley chromosome arm 1HL into a 637 kb region from which 13 co-segregated markers were developed with an 1890-line NIL-derived population.

2.4 Pan-genomics in crops

Phenotypic diversity is largely determined by a spectrum of genomic variations including single nucleotide polymorphism (SNP), small insertion/deletion (InDel), copy number variation (CNV) and presence/absence variation (PAV) (Gabur et al. 2019). In the past two decade, SNP and small InDel between individuals are primarily targeted by comparative genomic studies as their detection required only one reference genome and low-cost re-sequencing data of different genotypes. However, it has been increasingly noticed that a single reference genome and unassembled resequencing sequence are not enough for identifying structural variations (Saxena et al. 2014) including CNV and PAV which are prevalent in crops and significantly influence agronomic trait (Hirsch et al. 2014; Li et al. 2014; Lu et al. 2015; Springer et al. 2009; Zhao et al. 2018) and disease resistance (Gabur et al. 2019; Rawat et al. 2016; Su et al. 2019b). Therefore, a pan-genome containing full genetic elements from a species is necessary to thoroughly capture all genetic variations within the biological clade. The pan-genome concept was initially proposed in bacteria *Streptococcus agalactiae* to describe the core genes that shared by all strains and dispensable genes which are present in partial or individual strains (Tettelin et al. 2005). In the past several years, the pan-genomic

study has been increasingly popular in higher plants and different strategies for pan-genome construction have been developed.

2.4.1 Strategies of pan-genome construction

The straightforward approach to construct pan-genome, as the method applied in bacteria, is generating *de novo* assemblies of multiple genotypes or strains for the species with small and simple genome (Gordon et al. 2017; Tettelin et al. 2005; Zhao et al. 2018). For the plants with huge and highly repetitive genome, however, it is unfeasible to *de novo* sequencing and assembling numerous genomes for those organisms in the foreseeable future due to the financial burden and laborious computational cost. As an alternative, another approach aims to construct an in-silico representation of pan-genome by iteratively integrating the novel sequence found in following assemblies into the previous pan-genome reference. Though such a strategy can capture CNV and PAV cost-efficiently (Golicz et al. 2016; Montenegro et al. 2017; Zhou et al. 2017), it would entail an ascertainment bias due to the copy numbers of some sequence were predominantly determined by the *de novo* assembled genotypes (Monat et al. 2019b).

2.4.2 Progress of pan-genome construction in model organism and crops

2.4.2.1 *Brachypodium*

Bachypodum distachyon, the model organism for grass research (Brkljacic et al. 2011; Draper et al. 2001), has a small genome (~350 Mb) and rapid reproduction cycle. So far, the *Bachypodum* pan-genome consists of 54 *de novo* assemblies which are sequenced from Illumina short-read platform with an average 92X read depth and an average N50 of 75 kb (Gordon et al. 2017). Adding the previous reference genome (International Brachypodium Initiative 2010) in, the 55 *Brachypodium* genomes were iteratively integrated into a ~430 Mb *in silico* pan-genome representation (Gordon et al. 2017). The core genes account 73% of all genes annotated from the pan-genome and they significantly gathered in the pericentromeric regions.

2.4.2.2 Rice

The studies of rice pan-genome are proceeded much faster than other major crops (maize, wheat, barley) due to its relatively small genome (~400 Mb). Xu et al. (2012) reported the first rice pan-genome consists of 50 wild and domesticated genotypes with an average whole-genome depth of 15X Illumina short-reads. From this pan-genome, 1415 reference-absent predicted gene model that are orthologous with other plant species and most of them are present in no more than five genotypes. Wang et al. (2018a) reported the main outcome of the 3,000 Rice Genome Project which re-sequenced 3,010 diverse Asian cultivated rice on the Illumina platform. This study detected more than 90,000 structural variations from the huge population and identified over 10,000 novel protein-coding genes. As aligning short-reads into a single reference could often lose many genetic variations, Zhao et al. (2018) deep sequenced (average depth 115X) and *de novo* assembled 66 genotypes selected as representatives for 1579 wild and domesticated rice accessions (Huang et al. 2012). The comparative analysis identified over 23 million genetic variants from the rice pan-genome and 10,872 genes were entire or partially absent from the Nipponbare reference.

2.4.2.3 Maize

The maize has a 2.3-Gb genome enriched with highly repetitive sequences. The recombination rates along maize chromosomes are lower at peri-centromeric regions and higher at the distal regions (Schnable et al. 2009). SanMiguel et al. (1996) firstly reported maize genome is full of retrotransposons in the intergenic regions through single-locus analysis of *de novo* assembled contigs generated by bacterial artificial chromosomes (BACs). With the development of sequencing and assembly technology, seventeen genome assemblies (including updated versions) of 12 maize genotypes have been released in the past decade (Table 2.1).

Table 2.1 Overview of the genome assemblies for maize

Assembly Name	Line	Accession	Quality	Release
B73 RefGen_v1	B73	PRJNA10769	Representative	2011
B73 RefGen_v2	B73	PRJNA10769	Representative	2010
B73 RefGen_v3	B73	PRJNA72137	Representative	2013
Zm-B73-REFERENCE-GRAMENE-4.0	B73	-	Representative	2017
Zm-B104-DRAFT-ISU_USDA-0.1	B104	-	Draft	2017
Zm-CML247-DRAFT-PANZEA-1.0	CML247	-	Draft	2016
Zm-CML247-REFERENCE-PANZEA-1.1	CML247	PRJNA396542	Reference	2017
Zm-DK105-REFERENCE-TUM-1.0	DK105	PRJNA360923	Reference	2018
Zm-EP1-REFERENCE-TUM-1.0	EP1	PRJNA360920	Reference	2017
Zm-F7-REFERENCE-TUM-1.0	F7	PRJNA360923	Reference	2017
Zm-Mo17-REFERENCE-YAN-1.0	Mo17	PRJNA299869	Reference	2017
Zm-Mo17-REFERENCE-CAU-1.0	Mo17	PRJNA358298	Reference	2018
ZeaMays_PT_EDMX2233_1.0	Palomero Toluqueno	PRJNA51041	Scaffolds	2011
Zm-PE0075-REFERENCE-TUM-1.0	PE0075	PRJNA360923	Reference	2018
Zm-PH207-REFERENCE_NS-UIUC_UMN-1.0	PH207	PRJNA389728	Reference	2016
Zx-PI566673-REFERENCE-YAN-1.0	PI 566673	PRJNA299874	Reference	2017
Zm-W22-REFERENCE-NRGENE-2.0	W22	PRJNA311133	Reference	2016

Source: https://www.maizegdb.org/genome/assemblies_overview

Due to the huge size and complexity, however, it is still costly and laborious to sequence and assembly enough number of maize genomes for constructing a maize pan-genome. Instead, Hirsch et al. (2014) conduct the initial study on the pan-transcriptome of maize through RNA-sequencing seedling tissue of 503 maize inbred line. This study identified over 8,600 non-reference transcripts and most of them were successfully allocated to a relative position on reference genome by association mapping. Lu et al. (2015) developed a maize pan-genome consists of 4.4 million short sequence tags generated from 14,129 genotypes as proxies of pan-genomic fragments. One-quarter of those short sequences were identified as PAV tags. These tags were found have significant influences on agronomic traits including days to silk, days to anthesis, plant height and ear height.

2.4.3 Strategies for constructing barley Pan-genome

As the barley genome is much larger than maize, the pan-genomic study on barley has just emerged recently and started from pan-transcriptome as well as maize. Through *de novo* assembling 288 sets of RNA-seq data from 32 cultivated genotypes and 31 wild genotypes, Ma et al. (2019) established a barley pan-transcriptome including over 750,000 transcripts of which 38.2% are absent from the international reference genome of cultivar Morex (Mascher et al. 2017). This study also found that the wild genotypes enriched more disease resistance genes than cultivars.

In terms of constructing barley pan-genome with enough genotypes, the two major concerns are still the laborious assembly procedures and expensive sequencing cost. The assembly algorithm combining deep-coverage short-reads and complementary mapping datasets need 6 months to establish one chromosome-scale barley or wheat genome (Schreiber et al. 2018). Monat et al. (2019a) developed an open-source workflow, TRITEX, which can generate pseudomolecule-scale genome with paired-end, mate-pair, chromosome conformation capture sequencing data. Though this workflow is able to produce a chromosome-scale barley genome in no more than 4 weeks (all required input data are in place), building dozens or even hundreds of barley genomes is still time-consuming. In addition, the current cost of obtaining a chromosome-scale barley genome through *de novo* assembling only Illumina short-reads (Monat et al. 2019b) is about AUD 100,000. Such a high cost further inhibits the construction of a large number of barley genotypes in the immediate future. Therefore, the on-going barley pan-genome project (Monat et al. 2019b) aims to generate only a dozen of pseudomolecule-

scale references as a basis and then sequence hundreds to thousands of genotypes with medium coverage.

Unlike the construction of pseudomolecule-scale assembly, genotype-by-sequencing (GBS), a restriction enzyme-based reducing genome sequencing technique (Poland et al. 2012), is cost-effective and, therefore, feasible for applying on huge numbers of genotypes. The GBS dataset of all barley accessions in German ex situ genebank, comprising 21,405 accessions, has been released with relevant phenotypic data recently (Milner et al. 2019). Considering the GBS-based pan-genome in maize (Lu et al. 2015), it is feasible and cost-efficient to establish a partially informative barley pan-genome through analyzing the short sequences from the barley GBS dataset using genetic mapping and machine learning.

2.5 Summary

Barley is the fourth most important crop worldwide and FCR is one of the devastating diseases affecting its production. This disease, caused by various species of *Fusarium*, can not only severely decrease the yield in barley, but also produce mycotoxins in all tissues including grains, which is harmful to health of human and livestock. Even though some farming practices were used to control FCR, these practices may not always be economical, practical and friendly for the environments. It is well recognized that breeding and growing resistant varieties forms a critical component in mitigating the damage from this disease, for which the identification of genetic resources and resistance genes is critical. Previous studies only reported three major QTL conferring FCR resistance in barley thus there was an urgent need to identify additional resources of FCR resistance for effectively breeding resistant varieties. Due to limited resolution, markers obtained from QTL mapping are not necessarily tightly linked with the targeted gene. Fine mapping of these QTL is needed to develop new markers that can be reliably used in breeding programs. In addition, an individual genotype contains only a proportion of the genes found in a given species. A pan-genome consisting of genes from a large number of genotypes for a single species would form a solid foundation for efficient genetic studies and breeding in the post-sequencing era. With the rapid progress in sequencing technology and capacity, it has been become feasible to construct a pan-genome for barley which has a huge and complex genome.

Chapter 3 A novel QTL conferring *Fusarium* crown rot resistance located on chromosome arm 6HL in barley[1]

[1] This paper has been published as: Gao, S., Zheng, Z., Hu, H., Shi, H., Ma, J., Liu, Y., Wei, Y., Zheng, Y., Zhou, M. and Liu, C., 2019. A novel QTL conferring Fusarium crown rot resistance located on chromosome arm 6HL in barley. Frontiers in Plant Science, 10: 1206.

3.1 Introduction

Fusarium crown rot (FCR) is a chronic disease caused by various *Fusarium* species in cereal crops. It is prevalent in arid and semi-arid cropping regions worldwide (Akinsanmi et al. 2004; Chakraborty et al. 2006). The disease has been known for causing significant yield loss in Australia (Murray and Brennan 2010; Murray and Brennan 2009) and the USA (Smiley et al. 2005b). FCR has also become a major issue in recent years for cereal production in China (Li et al. 2016; Xu et al. 2017; Zhou et al. 2014).

FCR pathogens are carried over in crop residues. They can survive for two seasons or longer in the field environments (Chakraborty et al. 2006; Smiley et al. 2005a) making the disease difficult to manage using practices such as crop rotation (Burgess 2014). As recognised several decades ago, growing resistant varieties has to be an integral component in effectively managing the disease (Purss 1966; Wildermuth and Purss 1971). Growing resistant varieties does not only reduce yield losses but could also reduce the loss of the following barley or other cereal crops by reducing the inoculum load. This is especially the case for barley as, compared with those of wheat, its plants accumulate much higher concentrations of *Fusarium* pathogens at every stage of CR infection (Liu et al. 2012a).

The availability of well-characterized genotypes with high levels of resistance would facilitate breeding varieties with enhanced resistance. To date, there are only three reported studies on identifying QTL conferring FCR resistance in barley. The first one was based on a study using an existing population developed for unrelated traits. The study detected a single locus on chromosome arm 3HL which interacts strongly with plant height (Li et al. 2009). Effects of plant height on FCR development were also detected based on near isogenic lines (Chen et al. 2014) and histological analyses (Bai and Liu 2015). The second study detected a single locus on chromosome arm 4HL (designated as *Qcrs.cpi-4H)* from a wild barley genotype (*Hordeum spontaneum* L.) originated from Iran (Chen et al. 2013a). The third study detected two loci

from a landrace originated from Japan. They located on 1HL (*Qcrs.cpi-1H)* and 3HL (*Qcrs.cpi-3H*), respectively (Chen et al. 2013b).

Gene pyramiding has shown to be effective in further improving FCR resistance in both barley (Chen et al. 2015) and wheat (Zheng et al. 2017). However, as discussed above, only three loci conferring FCR resistance have been reported in barley. The value of one of them is also questionable as it interacts strongly with plant height (Chen et al. 2014; Li et al. 2009). With the aim of identifying additional loci with high levels of resistance to FCR, we conducted a QTL mapping study against AWCS799 which was a landrace originated from South Korea. This genotype was identified as one of the most resistant genotypes from a systematic screening of more than 1,000 genotypes representing diverse geographical origins and different plant types (Liu et al. 2012b). As previous studies have repeatedly shown that both plant height (Bai and Liu 2015; Chen et al. 2013a; Li et al. 2009; Ma et al. 2010; Zheng et al. 2014) and heading date (Liu et al. 2012b) may interact with FCR severity, we investigated possible interactions between QTL detected from this study with these characteristics. Results obtained from these analyses are reported in this publication.

3.2 Materials and Methods

3.2.1 Plant materials

The genotype AWCS799 is the resistant source analysed in this study. Two populations of recombinant inbred lines (RILs) between 'AWCS799' and two cultivars, Fleet and Franklin, were developed and used in this study. They are:

1. Fleet/AWCS799 consisting of 124 F8 RILs.

2. Franklin/AWCS799 consisting of 121 F8 RILs.

Both populations were produced in glasshouses at the Queensland Bioscience Precinct (QBP) in Brisbane, Australia. The first population was used for QTL mapping and the other for validating putative QTL identified from the mapping population.

3.2.2 FCR inoculation and disease assessments

Results from previous studies show that FCR resistance is not pathogen species-specific and the same resistance locus can be detected by pathogen isolates belonging to different *Fusarium* species (Chen et al. 2013b; Ma et al. 2010). We therefore used a single *F. pseudograminearum* isolate in this study. The isolate (CS3096) was obtained from a wheat field in northern New South Wales, Australia and maintained in the CSIRO collection (Akinsanmi et al. 2004). Inoculum preparation, inoculations, and FCR assessments were as described by Li et al. (2008). Briefly, inoculum was prepared using plates of ½ strength potato dextrose agar. Inoculated plates were kept for 12 days at room temperature before the mycelia were scraped and discarded. The plates were then incubated for a further 7-12 days under a combination of cool white and black fluorescent lights with 12-h photoperiod. The spores were then harvested, and the concentration of spore suspension was adjusted to 1×10^6 spores/ml. The spore suspensions were stored in minus 20 freezer and Tween 20 was added (0.1%v/v) to the spore suspension prior to use.

Seeds were germinated in Petri dishes on three layers of filter paper saturated with water. Seedlings of 3-day-old were immersed in the spore suspension for 1 min and two seedlings were planted into a 3 cm square punnet (Rite Grow Kwik Pots, Garden City Plastics, Australia) containing sterilized University of California mix C (50% sand and 50 % peat v/v). The punnets were arranged in a randomized block design and placed in a controlled environment facility (CEF). Settings for the CEF were: 25/18 (±1) °C day/night temperature and 65/80% (±5)% day/night relative humidity, and a 14-h photoperiod with 500 µmol $m^{-2}s^{-1}$ photon flux density at the level of the plant canopy. To promote FCR development, water-stress was applied during plant growth. Inoculated seedlings were watered only when wilt symptoms appeared.

For QTL mapping, four independent FCR severity tests were carried out against the mapping population (designated as FCR01 to FCR04, respectively). Three independent tests were conducted on the validation population (designated as FCRV01, FCRV02, FCRV03, respectively). Fourteen seedlings for each of the RILs and both parental genotypes were used for each of the tests. FCR severity was assessed 4 weeks after inoculation, using a 0 (no visible symptom) to 5 (whole plant severely to completely necrotic) scale as described by Li et al. (2008). Mean of scores for each line was used as disease index (DI) in QTL analysis.

3.2.3 Evaluation for plant height and heading date

To assess possible effects of plant height (PH) and heading date (HD) on FCR resistance, a trial consisting of three replicates was conducted on the mapping population (Fleet/AWCS799) with randomized block design at the CSIRO Research Station at Gatton, Queensland (27°34′S, 152°20′E). For each replicate, 20 seeds for each of the RILF8 lines were sown in a single 1.5 metre row with a 25 cm row-spacing. PH was measured at maturity as the height from the soil surface to the tip of the spike (awns excluded). Six measurements were taken from the six tallest tillers in each row and the average of each line was used for statistical analyses. Heading date was recorded on the day when 50% of all plant were at Zadock's stage Z55 (Tottman et al. 1979).

3.2.4 Molecular marker analysis

Genotypes for the two parents and 94 RILs from the population Fleet/AWCS799 were generated by the Department of Economic Development, Jobs, Transport and Resources (DEDJTR), Victoria, Australia, according to a tGBS (an optimized approach for genotyping-by-sequencing) pipeline (Ott et al. 2017). SSR markers were then developed for putative QTL regions and used to genotype the whole mapping population. SSR-finder (https://github.com/GouXiangJian/SSR_finder) was used to screen the variants within QTL regions between the pseudomolecule of 'Morex' (Mascher et al. 2017) and an assembly of a wild barley (*Hordeum spontaneum*) genotype *AWCS276* (Liu et al. 2020). Primers were designed using Primer-BLAST (Ye et al. 2012).

PCR reactions for the amplification of the SSR markers were carried out in a total volume of 12 µl containing 25 ng genomic DNA, 0.2 µM of forward and reverse primer, 3 mM $MgCL_2$, 0.2 mM dNTPs, and 0.5 U *Taq* DNA polymerase. During PCR, marker products were labelled with α-[^{33}P]dCTP (3,000 ci/mmol). PCR reactions were run on a Gene Amp PCR system 2700 thermocycler (PE Applied Biosystems, Foster City, CA, USA) programmed with the cycling conditions: one cycle of 5 min at 94 °C, 35 cycles of 30 sec at 94 °C, 30 sec at 60 °C and 1 min at 72 °C, with a final extension step of 5 min at 72 °C. The amplified products were mixed with an equal volume of loading dye, denatured at 95 °C for 5 min, and 4 µl samples was run on a denaturing 5% polyacrylamide (20:1) gel at 110 W for 2 h. The gels were subsequently dried using a gel dryer for 30 min at 80 °C and exposed to Kodak X-Omat X-ray films for 2 days.

3.2.5 Data analysis and QTL mapping

To generate the best linear unbiased prediction (BLUP) for DI values from the four FCR severity tests, R package 'lme4' (Bates et al. 2007) was used with the following mixed-effect model: $Y_{ij} = \mu + r_i + g_j + w_{ij}$. Where: Y_{ij} = DI value on the *jth* genotype in the *ith* test; μ = general mean; r_i = effect due to *ith* test; g_j = effect due to the *jth* genotype; w_{ij} = error or genotype by test interaction, where genotype was treated as a random effect and that of test as fixed. The graph for the frequency distribution of DI values and Pearson correlation coefficient between all phenotypic data were generated with Microsoft Office Excel 2016 and R package 'psych' (Revelle 2017), respectively. Student's *t* test was performed to evaluate the difference in DI values between lines with or without the resistant allele in the populations using Microsoft Office Excel 2016.

MSTmap Online (Wu et al. 2008) was used to build linkage maps by chromosomes with the following parameters: Grouping LOD Criteria: single LG; Population type: RIL8; No mapping missing threshold: 10%; No mapping distance threshold: 15 cM; No mapping size threshold: 2; Try to detect genotyping errors: Yes; Genetic mapping function: Kosambi. The diagrams of linkage maps were generated with MapChart (Voorrips 2002). IciMapping 4.1 was used for QTL analysis with the "Biparental Population" (BIP) module (Meng et al. 2015). Inclusive composite interval mapping (ICIM) was applied to identify QTL with a mapping step of 1cM (PIN = 0.015). For each FCR severity test, a 1,000-permutation test was performed to decide the LOD threshold corresponding to a genome-wide type I error less than 5% ($P < 0.05$). QTL were named according to the International Rules of Genetic Nomenclature (http://wheat.pw.usda.gov/ggpages/wgc/98/Intro.htm).

3.3 Result

3.3.1 Characterization of FCR resistance in the mapping population of Fleet/AWCS799

FCR severity of the resistant genotype AWCS799 was significantly lower than the two susceptible commercial cultivars (Fig. 3.1). In the four FCR severity tests conducted against the mapping population and its two parents, the DI value of AWCS799 was 50.0% lower than that of Fleet on average and transgressive segregation was detected in each of the tests (Table

3.1). DI values for all four tests and BLUP presented significantly positive correlation between each other (Table 3.2). The frequency distribution of DI value for FCR01 skewed towards better resistance. DI values for the other three tests and BLUP showed more normal distributions (Fig. S1.1).

Table 3.1 Disease index of FCR severity in the population of Fleet/AWCS799

Test	Parent		Population			
	Fleet	AWCS799	Min	Max	Mean	SD [a]
FCR01	3.4	1.3	0.9	4.2	2.2	0.9
FCR02	3.4	1.5	1.1	4.7	2.9	0.8
FCR03	3.6	2.4	1.1	4.6	2.7	0.8
FCR04	4.2	2.1	1.4	4.4	3.0	0.7
BLUP	3.4	1.6	1.2	4.1	2.5	0.6

[a] 'SD' stands for standard deviation.

Table 3.2 Correlation coefficients between FCR severity, plant height and heading date in the Fleet/AWCS799 population

	FCR01 [a]	FCR02	FCR03	FCR04	BLUP	PH	HD
FCR01	1.00						
FCR02	0.23**	1.00					
FCR03	0.88*	0.29**	1.00				
FCR04	0.33**	0.25**	0.36**	1.00			
BLUP	0.79**	0.56*	0.83**	0.66**	1.00		
PH [b]	-0.11	-0.10	-0.06	-0.10	-0.13	1.00	
HD [c]	-0.10	-0.11	0.03	0.06	-0.08	0.35**	1.00

[a] '*' significant at $p < 0.05$; '**' significant at $p < 0.01$.
[b] 'PH' stands for plant height.
[c] 'HD' stands for heading date.

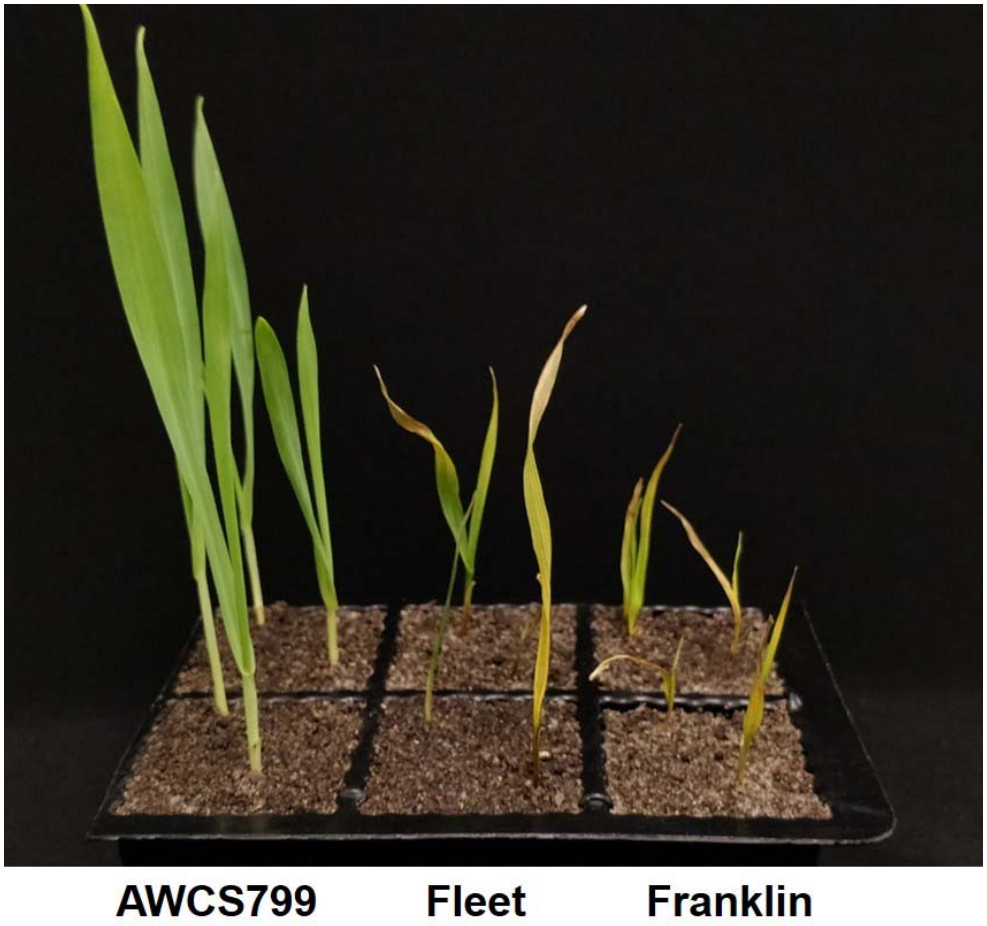


Fig. 3.1 Difference in resistance to *Fusarium* crown rot infection between the resistant genotype AWCS799 and the two commercial cultivars (Fleet and Franklin) used as parents in this study.

3.3.2 Linkage maps constructed and synteny for marker locations in the genome assembly of 'Morex'

Of the GBS markers mapped, 4,870 codominant markers were polymorphic between Fleet and AWCS799. These markers fell into 740 clusters and markers within each of the clusters co-segregated. As co-segregating markers contain the same information when used for mapping, a single marker with the least missing values was selected from each of the clusters and used for linkage map construction.

The markers were grouped into seven linkage groups and they spanned a total of 1964.7 cM with an average distance of 2.3 cM between loci (Table S3.1, for detail of linkage map, see Table S2). As all the markers generated had known physical positions, we aligned the linkage

maps with the reference genome of barley based on the cultivar 'Morex' (Mascher et al. 2017). This analysis found that, as expected, the genetic and physical maps were highly consistent for majority of the markers. However, there are a few exceptions around the peri-centromeric regions (Fig. S3.2).

3.3.3 Detection and validation of QTL for FCR resistance

Putative QTL were detected on chromosomes 1H, 2H, 3H, 4H, 5H and 6H. Only the one located on 6H was detected in each of the four tests conducted. The resistant allele of this QTL originated from AWCS799. We designated this QTL as *Qcrs.caf-6H*, where 'crs' stands for 'crown rot severity' and 'caf' represents 'CSIRO Agriculture and Food'. Qcrs.*caf-6H* was identified in all four tests and BLUP and it explained up to 28.3% of the phenotypic variance (Fig. 3.2 and Table 3.3). *Qcrs.caf-6H* was delineated into a 7.0 cM-interval flanked by markers 6H_453483214 and 6H_481998837 with 6H_497772849 (Table S3.3) as the most closely linked SSR marker.

Three QTL were identified on chromosome 1H. Of them, only *Qcrs.caf-1H.1* was detected in more than one test and it was located in a similar genomic interval with another FCR QTL reported in a previous study (Chen et al. 2013b). Four QTL were detected on chromosome 4H. Of them, *Qcrs.caf-4H.1* and *Qcrs.caf-4H.2* were identified in both FCR04 and BLUP. *Qcrs.caf-4H.2* was mapped in a similar location with a locus reported by Chen et al. (2013a). The other two 4H QTL and loci on chromosome 2H (*Qcrs.caf-2H),* 3H (*Qcrs.caf-3H*) and 5H (*Qcrs.caf-5H.1 and Qcrs.caf-5H.2*) were identified in only one of the tests. As none of these loci were consistently detected, they were not further investigated in this study.

Possible effects of *Qcrs.caf-6H* was further assessed in the validation population of Franklin/AWCS799. The most closely linked SSR marker with *Qcrs.caf-6H* from the mapping population was used to identify individuals with either the resistant (*RR*) or susceptible (*rr*) allele in this population. Significant difference was detected for *Qcrs.caf-6H* between *RR* and *rr* group in each of the three tests (Table 3.4). The average DI value of the lines bearing the resistant allele was 22.9% lower than those with the susceptible allele.

Table 3.3 Results of QTL analysis for FCR severity, plant height and heading date identified in the population of Fleet/AWCS799

Tests [a]	QTL	Interval	Flanking markers	LOD	PVE(%)[b]	Origin
BLUP	*Qcrs.caf-6H*	118.5 - 125.5	6H_453483214 & 6H_481998837	14.6	28.3	AWCS799
	Qcrs.caf-4H.1	195.5 - 197.5	4H_613436155 & 4H_615703492	3.1	4.5	Fleet
	Qcrs.caf-4H.2	208.5 - 210.5	4H_630512237 & 4H_633453171	7.0	11.4	AWCS799
	Qcrs.caf-2H	194.5 - 197.5	2H_638031213 & 2H_639632684	3.0	4.4	AWCS799
	Qcrs.caf-1H.1	170.5 - 239.5	1H_488323557 & 1H_540988817	3.3	4.8	AWCS799
FCR01	*Qcrs.caf-6H*	118.5 - 125.5	6H_453483214 & 6H_481998837	9.8	12.7	AWCS799
	Qcrs.caf-5H.1	135.5 - 141.5	5H_410496043 & 5H_451983882	5.1	5.8	AWCS799
	Qcrs.caf-5H.2	144.5 - 146.5	5H_450803506 & 5H_459449531	10.3	13.5	Fleet
	Qcrs.caf-4H.3	47.5 - 51.5	4H_28747562 & 4H_32616055	3.4	3.7	AWCS799
	Qcrs.caf-1H.1	202.5 - 211.5	1H_520276315 & 1H_524761112	5.2	6.2	AWCS799
FCR02	*Qcrs.caf-6H*	118.5 - 125.5	6H_453483214 & 6H_481998837	4.6	15.9	AWCS799
	Qcrs.caf-5H	31.5 - 34.5	3H_26301091 & 3H_27911717	6.0	20.6	Fleet
FCR03	*Qcrs.caf-6H*	118.5 - 125.5	6H_453483214 & 6H_481998837	3.8	6.4	AWCS799
	Qcrs.caf-1H.2	81.5 - 86.5	1H_300306311 & 1H_134967717	4.7	8.0	Fleet
	Qcrs.caf-1H.3	95.5 - 98.5	1H_377585283 & 1H_358493268	8.2	15.1	AWCS799
FCR04	*Qcrs.caf-6H*	118.5 - 125.5	6H_453483214 & 6H_481998837	9.6	12.0	AWCS799
	Qcrs.caf-4H.4	92.5 - 95.5	4H_345503178 & 4H_362989846	8.9	10.8	AWCS799
	Qcrs.caf-4H.1	195.5 - 197.5	4H_613436155 & 4H_615703492	5.7	6.5	Fleet
	Qcrs.caf-4H.2	208.5 - 210.5	4H_630512237 & 4H_633453171	12.2	16.7	AWCS799
	Qcrs.caf-1H.1	170.5 - 239.5	1H_488323557 & 1H_540988817	3.0	3.3	AWCS799
PH	*Qcrs.ph-7H*	118.2 - 124.7	7H_180473897 & 7H_408998941	7.4	27.8	AWCS799
	Qcrs.ph-6H	160.5 - 163.3	6H_555603298 & 6H_552874297	3.7	11.7	Fleet
HD	*Qcrs.hd-5H*	248.4 - 252.4	5H_594490721 & 5H_599429072	14.4	48.9	AWCS799

[a] 'PH' stands for plant height; 'HD' stands for heading date.

[b] Percentage of phenotypic variance explained.

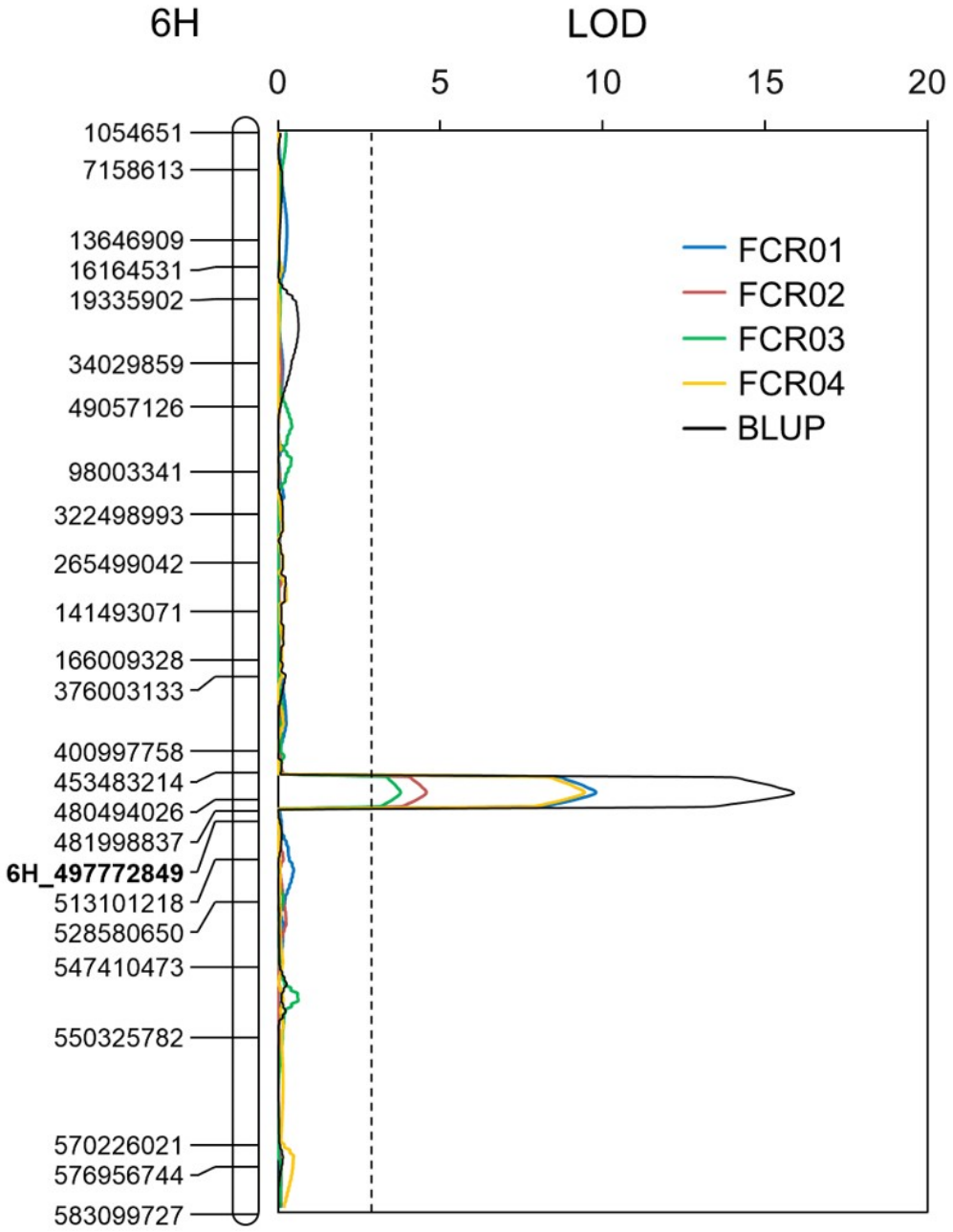


Fig. 3.2 QTL conferring FCR resistance detected on the long arm of chromosome 6H in the population of Fleet/AWCS799. Physical position for each of the markers is shown to the left of the linkage map. The vertical dotted lines indicate the average significance threshold (LOD=2.7) based on a test of 1,000 permutations. The SSR marker used in validating the QTL is in bold.

Table 3.4 Disease index of FCR severity of lines possess resistant (*RR*) and susceptible (*rr*) allele of *Qcrs.caf-6H* from the population of Franklin/AWCS799

Tests [a]	*RR* [b]	*rr* [b]	Difference (%) [c]	*P* value [d]
FCRV01	3.0	3.7	21.0	<0.01
FCRV02	2.2	2.7	18.7	<0.05
FCRV03	1.9	2.6	28.9	<0.05

[a] The three tests conducted were designated as FCRV01, FCRV02 and FCRV03, respectively.
[b] The number of *RR*- and *rr*- genotypes are 50 and 71.
[c] Differences were obtained by comparison between *RR*- and *rr*- genotypes.
[d] '*P* values' were generated with student's *t* test

3.3.4 Effect of plant height and heading date on FCR resistance

A QTL controlling HD was identified on chromosome 5H. This QTL explained up to 48.9% of phenotypic variance with a LOD value of 14.4 (Table 3.3). Two QTL affecting PH were detected and were located on chromosomes 6H and 7H, respectively. The QTL on chromosome 6H could explain up to 11.7% of the phenotypic variance with a LOD value of 3.7 and the other QTL on chromosome 7H could explain up to 27.8% of the phenotypic variance with a LOD value of 7.4 (Table 3.3). For quantifying possible effects of PH and HD on FCR severity, the BLUP of the four FCR severity tests was analysed against PH and HD data using covariance analysis. The results showed that both PH and HD have little effect on *Qcrs.caf-6H* (Fig. 3.3).

4 Discussion

In the study reported here, we investigated the genetics of FCR resistance on a barley landrace originating from South Korea. A novel QTL located on chromosome arm 6HL was detected from each of the four tests. This QTL, designated as *Qcrs.cpi-6H*, explained up to 28.3% of the phenotypic variance in the mapping population and reduced FCR severity by 22.9% in the validation population on average. This is the first locus conferring FCR resistance identified on this chromosome in barley.

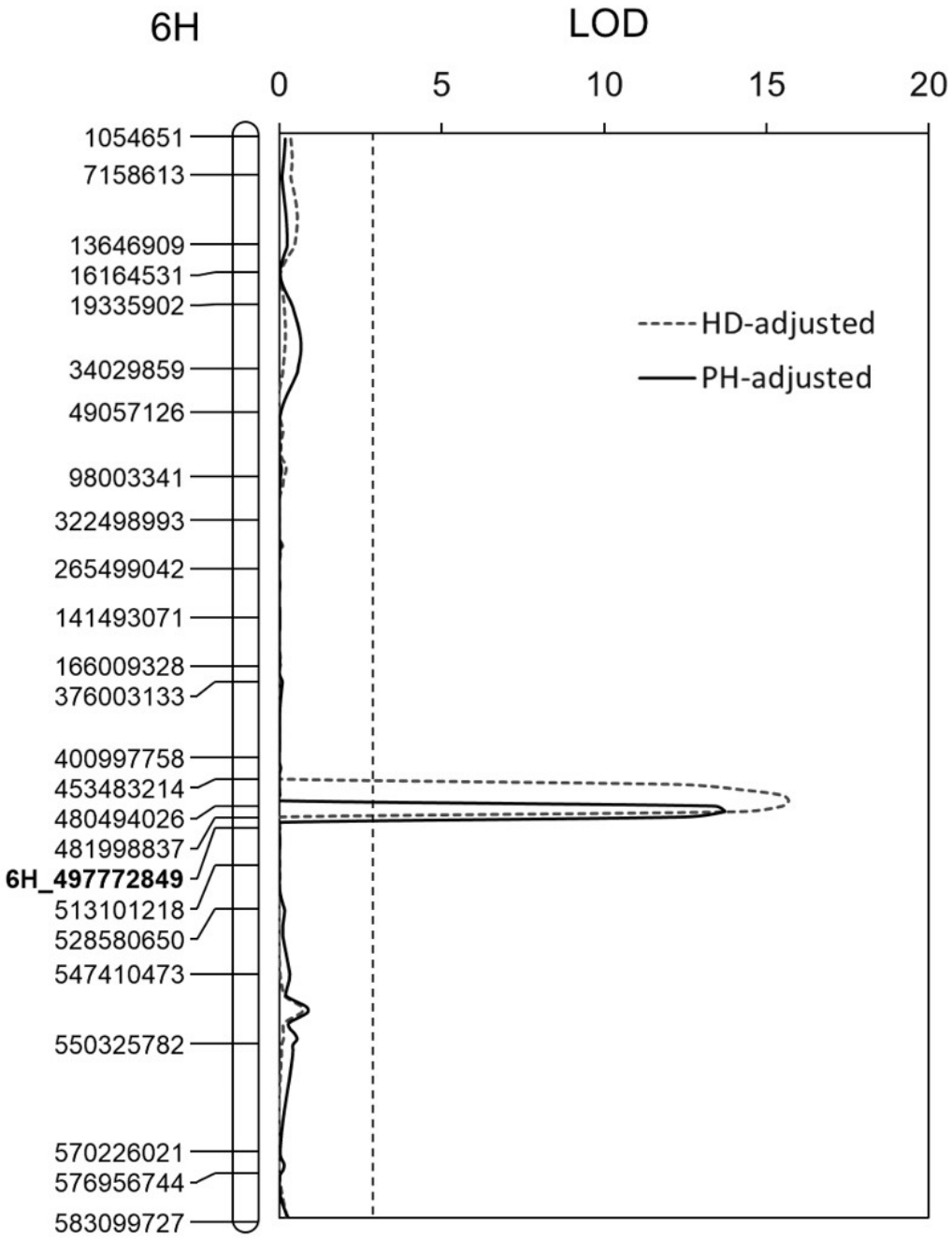


Fig. 3.3 Effects of plant height (PH) and heading date (HD) on Qcrs.caf-6H. Physical position for each of the markers is shown to the left of the linkage map. The LOD values were obtained from the BLUP and post-adjustment by HD and PH. The vertical dotted lines indicate the average significance threshold (LOD=2.7) based on a test of 1,000 permutations. The SSR marker used in validating the QTL is in bold.

Pyramiding multiple loci into single genetic background has been proved to be effective in improving FCR resistance in barley and wheat (Chen et al. 2015; Zheng et al. 2017). However, the progress of such work in barley has been hampered by the shortage of effective loci. Of the only three loci reported in earlier studies, the value of the one located on chromosome arm 3HL is questionable as it had strong interaction with plant height (Chen et al. 2013a; Li et al. 2009; Liu et al. 2012a). A histological analysis based on NILs for height also showed that *F. pseudograminearum* hyphae were detected earlier and proliferated more rapidly during the time-course of FCR development in the tall isolines (Bai and Liu 2015). Some of the earlier studies showed that FCR severity can also be influenced by heading date (Liu et al. 2012a). The new locus detected in this study does not co-locate with loci controlling either plant height or heading date and removing the effects of these characteristics from the mapping population has little influence on the magnitudes of the FCR locus detected in any of the tests. Thus, *Qcrs.caf-6H* is the third locus which can be effectively exploited to further enhance FCR resistance by gene pyramiding in this crop species.

Loci conferring FCR resistance have been reported on 13 of the 21 wheat chromosomes (Liu and Ogbonnaya 2015). One of the loci was located on wheat chromosome 6DL and it was detected from a field trial (Martin et al. 2015). Compared with the 6DL locus, *Qph.caf-6H* seems to be more distally located on the chromosome arm 6HL in barley. However, as QTL mapping provides only limited resolution (Paterson et al. 1988), further studies are required to clarify the homoeologous relationship between these two loci.

As expected, orders for most of the markers in the linkage map constructed in this study aligned well with their physical positions in the barley genome. It is of interest to note that, without any exception, all the discrepancies involved markers and sequences located in the peri-centromeric region (Fig. S3.2). It is known that the peri-centromeric regions of the barley chromosomes are characterized by low gene density, low recombination frequencies (Mascher et al. 2017) and high ratios of repetitive sequences (Wicker et al. 2017). Contributions, if any, from these characteristics to the discrepancies are not clear. However, it is not unreasonable to speculate that low recombination frequencies are likely to be less tolerable to incorrect marker scores or missing values occurred during high-throughput genotyping.

Chapter 4 Validation and delineation of a locus conferring *Fusarium* crown rot resistance on 1HL in barley by analysing transcriptomes from multiple pairs of near isogenic lines[2]

[2] This paper has been published as: Gao, S., Zheng, Z., Powell, J., Habib, A., Stiller, J., Zhou, M. and Liu, C., 2019. Validation and delineation of a locus conferring Fusarium crown rot resistance on 1HL in barley by analysing transcriptomes from multiple pairs of near isogenic lines. BMC Genomics, 20: 650.

4.1 Introduction

Fusarium crown rot (FCR), caused mainly by *F. pseudograminearum*, is a severe and chronic disease of cereals in semi-arid cropping regions worldwide (Chakraborty et al. 2006; Hogg et al. 2010). To reduce FCR damage, several agronomic measures have been developed. They include crop rotation and stubble management (Cook 1980, 2001). These practices can reduce the impact of FCR in certain circumstances but are not always useful due to economic and practical requirements (Kirkegaard et al. 2004). It has long been recognised that growing resistant varieties is an essential component to effectively manage this disease (Purss 1966).

The approach of identifying and transferring major QTL into elite genotype has been used in breeding FCR-resistant varieties in wheat and barley (Chen et al. 2015; Zheng et al. 2017). Up to date, four putative QTL conferring FCR resistance have been reported in barley (Liu and Ogbonnaya 2015). They locate on chromosome arms 1HL (Chen et al. 2013b), 3HL (Li et al. 2009), 4HL (Chen et al. 2013a) and 6HL (Gao et al. 2019b), respectively. Similar to those noticed in wheat (Yan et al. 2011; Zheng et al. 2015), strong interactions between FCR severity and other characteristics including flowering time (Chen et al. 2013a; Liu et al. 2012b) and plant height (Li et al. 2009; Liu et al. 2010) have also been detected in barley. The FCR resistance locus on chromosome arm 3HL in barley also co-locates with gene(s) controlling spike structure (Chen et al. 2012). Results from previous studies also showed that water availability affects FCR development (Liu and Liu 2016).

The interactions between FCR severity and other characteristics indicate that QTL detected through mapping can only be treated as putative. The effectiveness of a QTL detected from segregating populations needs to be validated. Near isogenic lines (NILs) have been used widely in validating QTL for various characteristics (Chen et al. 2014; Pumphrey et al. 2007b).

They were also used to validate QTL conferring resistance to FCR in cereals (Habib et al. 2016; Ma et al. 2012).

The main focus of transcriptomic analysis was to detect differentially expressed genes (DEGs) when the technique was initially introduced (Mortazavi et al. 2008; Wang et al. 2009). The analysis is now also widely used to uncover genetic markers for various purposes (Blencowe et al. 2009; Cavanagh et al. 2013). Combined with the use of NILs, distributions of variations detected from transcriptomic sequences have been exploited effectively in validating QTL and obtaining markers for fine mapping targeted loci (Habib et al. 2018; Jiang et al. 2019; Ma et al. 2014).

In the study reported here, NILs were developed and used to validate the QTL conferring FCR resistance on 1HL. Transcriptomic sequences were then obtained from three pairs of the NILs. Shared SNPs detected from the transcriptomic sequences among the NIL pairs were used to further delineate the QTL interval and identify candidate genes underlying the resistance locus on 1HL.

4.2 Materials and methods

4.2.1 Development of near isogenic lines

The heterogeneous inbred family method (Tuinstra et al. 1997), combined with the fast-generation technique (Zheng et al. 2013), was used to develop NILs targeting the 1HL locus (*Qcrs.cpi-1H*). Plants were raised in glasshouses at Queensland Bioscience Precinct in Brisbane, Australia. Heterozygous plants were identified from two segregating populations, 'Locker//AWCS079/AWCS276' and 'Commander//AWCS079/AWCS276', using the SSR marker *WMC1E8.* This marker was one of those linked closely with *Qcrs.cpi-1H* identified from QTL mapping (Chen et al. 2013b). Primer sequences of the marker were: forward 5'-TCATTCGTTGCAGATACACCAC-3'; and reverse 5'-TCAATGCCCTTGTTTCTGACCT-3'. The identified plants were self-pollinated for eight generations and a single pair of putative NILs was then selected from each of the original heterozygous plants.

4.2.2 FCR inoculation and assessment

FCR inoculation was conducted in the controlled environment facilities (CEFs) at Queensland Bioscience Precinct, Brisbane. Four independent trials were conducted against the putative

NILs. Each trial consists of two replicates and 14 seedlings per isoline were used in each of the replicates. A highly aggressive isolate of *F. pseudograminearum* (CS3096) was used for inoculation in these trials. This isolate was collected in northern New South Wales and maintained in the CSIRO collection (Akinsanmi et al. 2004). Procedures used for inoculum preparation, inoculation and FCR assessment were based on those described by Li et al. (Li et al. 2008). Briefly, seeds were surface-sterilized by treating with 2% hypochlorite solution for 10 min and then thoroughly rinsed with distilled water for four times. The seeds were then germinated on three layers of filter paper saturated with water in petri-dishes. Newly germinated seedlings (with coleoptile lengths ranging from 0.5 to 1.0 cm) were inoculated by immersing in *Fusarium* spore suspension (or water for controls) for 1 min. Two treated seedlings were sown in a 4cm x 4cm square punnet (Rite Grow Kwit Pots, Garden City Plastics, Australia) containing autoclaved potting mix. Fifty-six punnets were placed in a plastic seedling tray for easy handling. Inoculated seedlings were kept in CEFs. Settings for the CEFs were: 25/16(± 1) °C day/night temperature and 65%/85% day/night relative humidity, and a 14-h photoperiod with 500 mol m−2 s−1 photon flux density at the level of the plant canopy. Plants were watered only when wilt symptoms appeared. FCR severity for each plant was assessed with a 0-5 scale, where "0" standing for no symptom and "5" representing whole plant necrotic (Li et al. 2008). Disease indices (DI) was calculated for each line following the formula of $DI = (\sum_n X / 5N) \times 100$, of which, X is the scale value of each plant, n is the number of plants in the category, and N is the total number of plants assessed for each line. The difference between the isolines possessing the resistant and susceptible allele for each of the putative NIL pairs was assessed with the student *t* test.

4.2.3 RNA extraction and sequencing

Samples for RNA sequencing were obtained from three pairs of the NILs. Inoculation was conducted with either the *F. pseudograminearum* isolate (*Fp*-inoculation) or distilled water (mock) following the protocol described above. Three biological replications were conducted for every isolines. Each replication consisted of seven seedlings. Tissues for RNA extraction were collected by cutting the shoot bases (2 cm) at 4 days post inoculation (dpi) and snap-frozen in liquid nitrogen and kept at – 80 °C until processed. The time point for sampling was selected based on a previous study (Habib et al. 2018).

A total of 36 samples were obtained from the six isolines. Samples were crushed into fine powder and RNA extraction was conducted using an RNeasy plant mini kit (Qiagen, Hilden,

Germany) according to manufacturer's instructions (including DNase-I digestion). The yield and purity of RNA samples were measured using a Nanodrop-1000 Spectrophotometer. The integrity of all RNA samples was assessed by running the total RNA on 1% agarose gels. RNA sequencing was carried out by the Australian Genome Research Facility Ltd (Parkville, Victoria, Australia) and 100-bp paired-end reads were produced using the Illumina Hiseq-2000. All 36 RNA-seq libraries were run across four lanes of a HiSeq2000.

4.2.4 Transcriptomic analyses

Commands used for trimming raw data and analysing trimmed reads were described by Habib et al. (Habib et al. 2018). FastQC (version 0.11.2) was used as a preliminary check for PHRED scores. Raw reads were trimmed using the SolexaQA package (version 3.1.3) with a minimum PHRED quality value of 30 and minimum length of 70 bp. TopHat2 (version 2.0.13) (Trapnell et al. 2012) was used to map filtered reads to the barley cv. Morex genome (https://webblast.ipk-gatersleben.de/barley_ibsc/downloads/: 150831_barley_pseudomolecules) which is now widely used widely as the reference for barley (Mascher et al. 2017).

Differential gene expression analysis: Cufflinks (version 2.0.2) (Trapnell et al. 2012) was used to assemble the mapped reads. DEGs were identified with Cuffdiff from the Cufflinks tool package with high-confidence genes annotated in the 'Morex' genome. Fragments per kilobase of exon per million mapped reads (FPKM) was applied for each transcript to represent the normalized expression value. The fold change in gene expression was calculated according to the equation: Fold Change = $\log_2 (FPKM_A / FPKM_B)$.

Pairwise comparisons were conducted between different treatments for the same isoline (S^M_v_S^I and R^M_v_R^I) and between isolines under *Fp*-inoculation (S^I_v_R^I) or mock-inoculation (S^M_v_R^M) (Fig. 4.1). 'M' stands for 'mock-inoculation', 'I' for *Fp*-inoculation, 'S' for susceptible isolines, and 'R' resistant isolines. DEGs were determined with the adjusted *p*-value threshold of ≤ 0.05 and $\log_2$ fold change of ≥ 1 or ≤ -1 or 'inf' (where the FPKM value in one dataset is zero and the other is not). Venny 2.0 was used for Venn diagram analysis (Oliveros 2018).

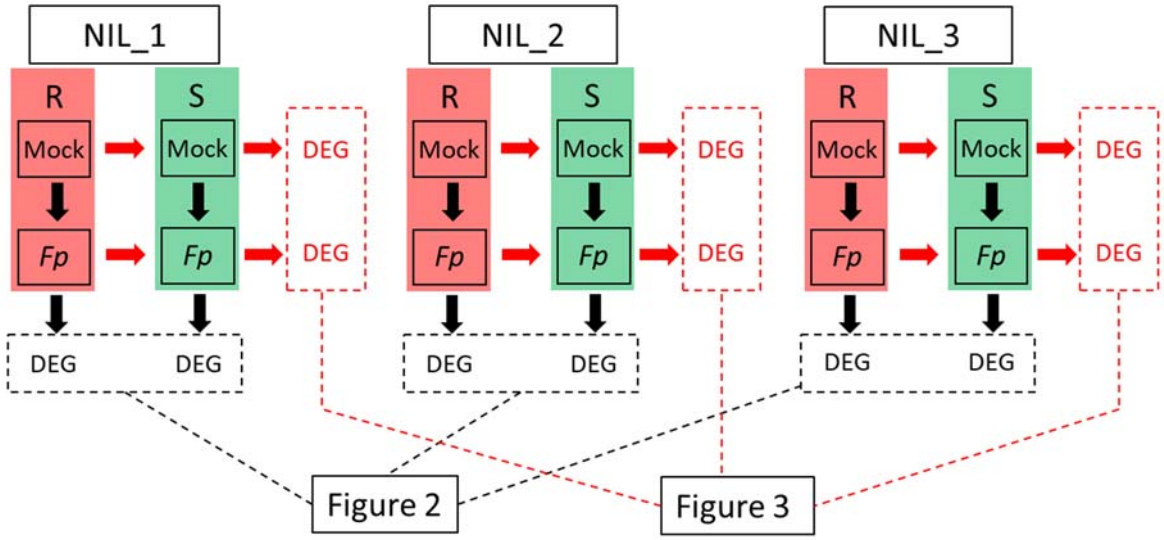


Fig. 4.1 The experimental design for differential gene expression analysis.

Validation of differentially expressed genes using qRT-PCR: Three genes (*HORVU1Hr1G092240*, *HORVU1Hr1G092250* and *HORVU1Hr1G092300*; primer sequences were listed in Table S4.1) were selected from the identified DEGs for validation. Quantitative real-time PCR (qRT-PCR) was used for validation with the actin protein gene as the internal housekeeping reference (forward primer: 5'-GCCGTGCTTTCCCTCTATG-3'; reverse primer 5′-GCTTCTCCTTGATGTCCCTTA-3′). Inoculation, tissue sampling and RNA extraction were carried out using the aforementioned methods. Three biological replicates, each with two technical replications, were used for each genotype-treatment sample per isoline.

The procedures for synthesising cDNA and qRT-PCR were conducted following the methods described by Ma et al. (2013). The relative fold changes were calculated using the comparative CT method ($2^{-\Delta\Delta CT}$). The average value of the two technical replications was used to represent the biological replicate for each of the samples.

SNP calling and nonsynonymous variation identification: For each genotype, all six sequence files (three biological replicates by two treatments) were concatenated after removing low-quality sequences. The concatenated files were then aligned to the 'Morex' genome using Biokanga align with a maximum of two mismatches per read. SNPs between the 'R' and 'S' isolines of each NIL pair were identified using the Biokanga snpmarkers with a minimum 80% score (the percentage of a given nucleotide at an SNP position is at least 80% in the 'R' or 'S'

isoline). The SNPs were annotated using snpEff 4.3q and the variant database was built based on the Morex genome and its annotation file (Mascher et al. 2017).

4.2.5 Gene annotation and GO term enrichment analysis

BLAST, mapping and annotation steps were performed using the standard parameters in BLAST2GO (Conesa et al. 2005) and the GO annotation results were used as reference (Table S4.6) in the following analysis. DEGs identified from all comparisons were separated into up-regulated and down-regulated gene lists (Table S4.6) and submitted to singular enrichment analysis using agriGO (Du et al. 2010; Tian et al. 2017) with default setting.

4.3 Results

4.3.1 Development and validation of NILs targeting the FCR resistance locus on 1HL

Eight heterozygous plants were initially selected from the two segregating populations based on the profiles of the SSR marker *WMC1E8*. A single pair of putative NILs was obtained from each of the heterozygous plants. Significant difference in morphology between any pairs of the putative 'R' and 'S' isolines was not observed. Significant difference in FCR severity was detected between the isolines for five of the eight putative NIL pairs. As expected, the isolines carrying the resistant allele from the donor parent AWC079 always gave much lower FCR severity than their counterparts (Table 4.1). The average DI for the 'R' isolines was 27.1, whereas it was 68.4 for the 'S' isolines. Three of the five NIL pairs with the largest difference in FCR severity, namely 1H_NILs: 1H_NIL1, 1H_NIL2 and 1H_NIL3, were selected and used for RNA-seq analysis.

Table 4.1 Difference in disease index between the resistant and susceptible isolines for the five NIL pairs targeting the 1HL locus conferring FCR resistance

NIL[a]	Genetic Background	DI Mean[b]	SE[c]	Difference (%)[d]	*P* value[e]
1H_NIL1_R	Lockyer//AWCS079/AWCS276 F8	24.9	4.2	66.1	<0.01
1H_NIL1_S		73.7	6.4		
1H_NIL2_R	Lockyer//AWCS079/AWCS276 F8	24.6	2.1	63.4	<0.01
1H_NIL2_S		67.3	4.1		
1H_NIL3_R	Commander//AWCS079/AWCS276 F8	26.4	1.8	58.0	<0.01
1H_NIL3_S		62.9	2.6		
1H_NIL4_R	Lockyer//AWCS079/AWCS276 F8	27.9	1.0	57.4	<0.01
1H_NIL4_S		65.5	1.4		
1H_NIL5_R	Commander//AWCS079/AWCS276 F8	31.7	2.5	56.4	<0.01
1H_NIL5_S		72.7	4.8		

[a] 'R' represents isolines with the allele from the resistant parent 'AWC079' and 'S' isolines with an alternative allele from the susceptible parents.
[b] The mean of disease index (DI value) observed from four trials for each isoline.
[c] 'SE' represents standard error.
[d] Differences between DI values of 'R' and 'S' isolines.
[e] '*P* value' was generated with the student's *t* test.

4.3.2 Transcriptome analyses

A total of 792 million quality reads were generate from the 36 samples (see the section of Materials and methods) with an average of 22 million reads per sample. The reads from each of the samples covered on average 21,571 high confidence (HC) genes (54.2% of all HC genes) based on the genome of Morex.

To analyse host response to *Fusarium* infection, DEGs were detected between *Fp*- (*F. pseudograminearum*-) and mock-inoculated samples of the same isoline. This analysis identified a total of 1,323 DEGs from the 'R' isolines and 2,083 from the 'S' isolines. The numbers of up-regulated genes were significantly higher than those down-regulated ones following *Fp*-inoculation (Table 2). Of the up-regulated genes, 144 were shared by all the three 'R' isolines and 370 by the three 'S' isolines (Fig. 4.1 and Fig. 4.2). Of the down-regulated genes, 17 were shared by the three 'R' lines and only 9 by the three 'S' lines. Expression patterns consistent with the RNA-seq analysis were obtained in the qRT-PCR analysis for each of the three genes assessed (Table S4.1).

Table 4.2 Number of differentially expressed genes (DEGs) identified from all pairwise comparisons

NIL pair	Comparison[a]	Number of DEGs	
		Up-regulated	Down-regulated
1H_NIL1	R^M_vs_R^I	226	60
	S^M_vs_S^I	831	113
1H_NIL2	R^M_vs_R^I	962	132
	S^M_vs_S^I	806	78
1H_NIL3	R^M_vs_R^I	910	117
	S^M_vs_S^I	1585	252
1H_NIL1	R^I_vs_S^I	48	236
	R^M_vs_S^M	225	123
1H_NIL2	R^I_vs_S^I	51	459
	R^M_vs_S^M	178	89
1H_NIL3	R^I_vs_S^I	249	132
	R^M_vs_S^M	80	71

[a]'M' stands for 'mock-inoculation', 'I' for *Fp*-inoculation, 'R' for resistant isolines and 'S' for susceptible isolines.

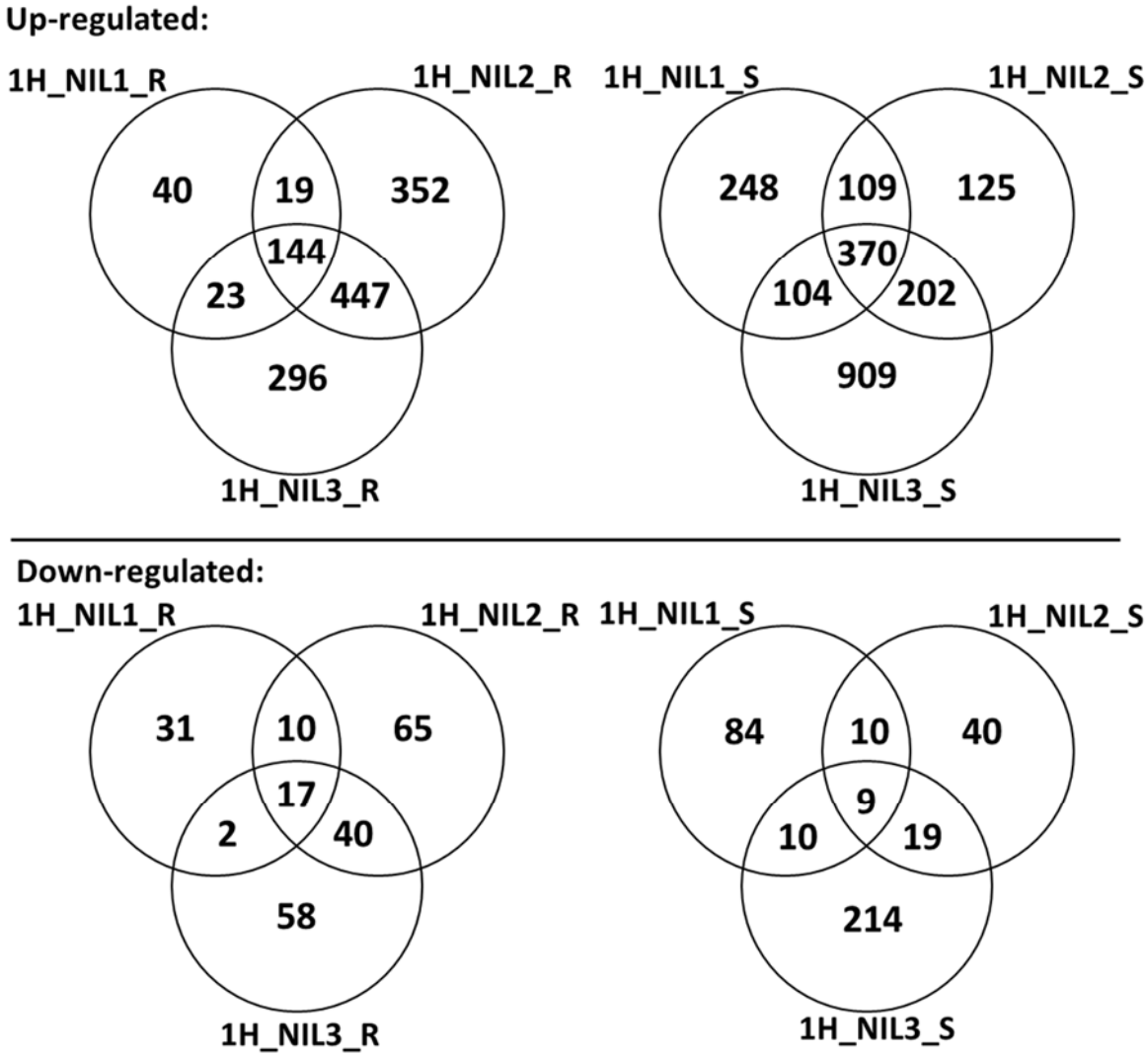


Fig. 4.2 DEGs for each of the 1H_NIL pairs following Fp- and mock-inoculation (RM_vs_RI and SM_vs_SI). Venn diagrams in upper panel show the numbers of up-regulated DEGs in each 'R' (left) and 'S' (right) isolines. Venn diagrams in lower panel show the numbers of down-regulated DEGs in each 'R' (left) and 'S' (right) isolines. DEGs were determined with the threshold of FDR ≤ 0.05 and |log2 fold-change|≥ 1 or 'inf' (one of the comparative objects did not express and the other did)

Gene ontology (GO) term enrichment analysis was performed on sets of differentially expressed genes from each comparison, separating out upregulated from downregulated genes. The goal of this approach was to isolate particular biological processes which might explain

the difference in resistance levels observed between 'R' and 'S' isolines. For genes up- or down-regulated during infection in 'R' isolines ('R^M vs R^I'), eleven, seventeen and twelve enriched terms were identified for 1H_NIL1_R, 1H_NIL2_R and 1H_NIL3_R respectively (Table S2). When observing genes upregulated during infection in susceptible isolines ('S^M vs S^I'), a total of six, nine and fifteen enriched terms were identified for 1H_NIL1_S, 1H_NIL2_S and 1H_NIL3_S respectively. Due to limited number of DEGs identified, no common enriched GO terms across pairwise comparisons for genes down-regulated during infection in 'R' or 'S' isolines were detected. GO term enrichment lists were compared to find terms commonly enriched across all three 'R' or 'S' isolines. For genes up-regulated during infection, three GO terms relating to the Cytochrome P450 superfamily (iron ion binding (GO:0005506), heme binding (GO:0020037) and tetrapyrrole binding (GO:0046906) were overrepresented consistently in both 'R' and 'S' isolines. In addition, glutathione transferase activity (GO: 0004364) was enriched across all three 'S' isolines and enriched across two 'R' isolines. GO terms enriched in only the three 'R' isolines or 'S' isolines were not detected. Results from the enrichment analysis inferred a common response to infection in both 'R' and 'S' isolines with terms having known roles in both biotic and abiotic stress responses. However, specific processes showing a consistent difference between 'R' and 'S' isolines which might explain increased resistance in 'R' isolines were not found at this relatively early infection timepoint.

To assess transcriptomic responses to FCR infection mediated by *Qcrs.cpi-1H*, we compared DEGs between the 'R' and 'S' isolines. These comparisons found that a total of 303 genes were up-regulated and 790 down-regulated from the *Fp*-inoculation treatment (Table 4.2). Only 4 of the up-regulated genes and 2 of the down-regulated ones were shared by all three NIL pairs (Fig. 4.1 and Fig. 4.3). Of the DEGs identified from the mock-inoculated samples, 440 were up-regulated and 283 down-regulated (Table 4.2). Ten of the up-regulated and 3 down-regulated ones were shared across all the three comparisons (Fig. 4.3).

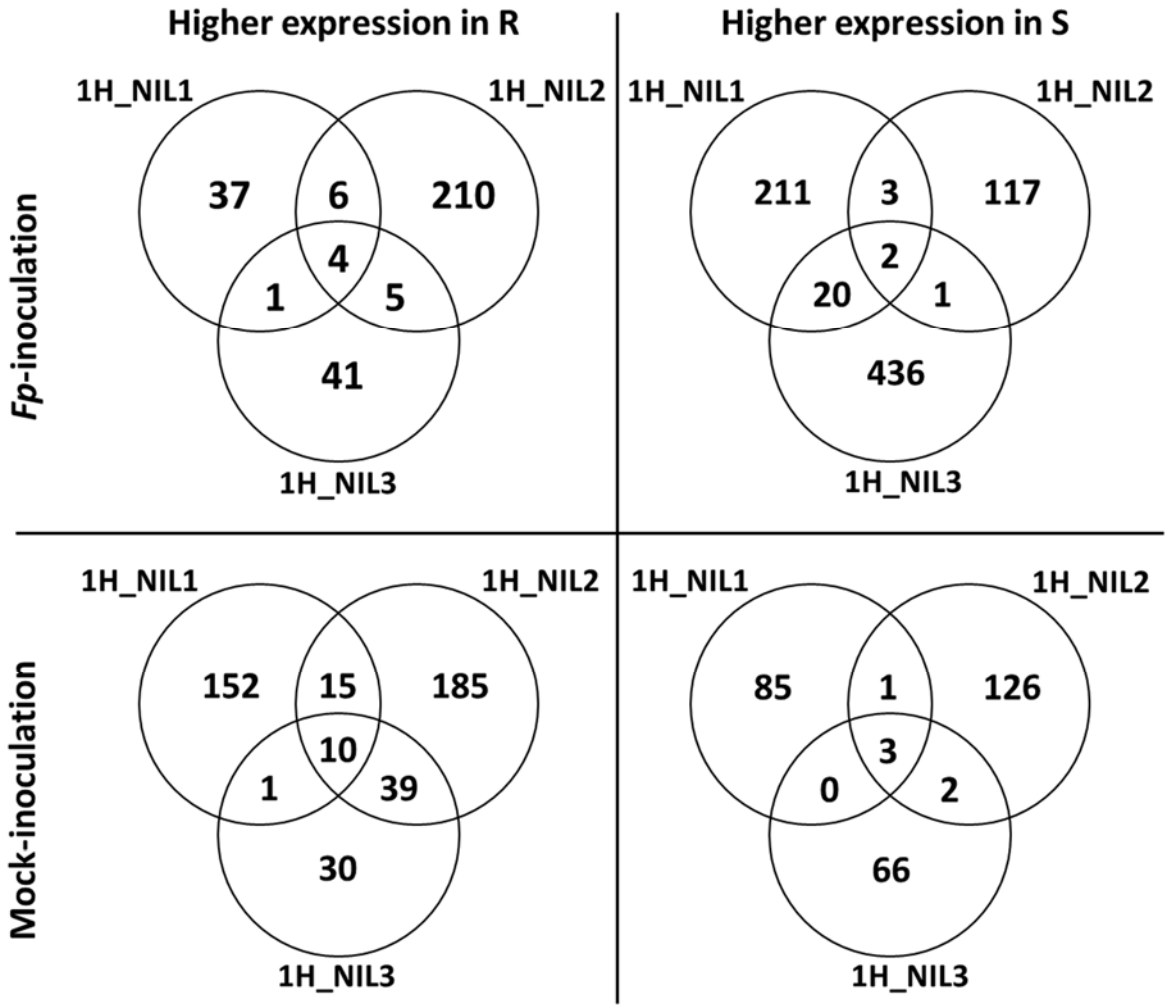


Fig. 4.3 DEGs between 'R' and 'S' isolines under Fp- (RI_vs_SI) or mock-inoculation (RM_vs_SM). Venn diagrams show the numbers of DEGs which up-regulated in 'R' (left) or 'S' (right) isolines under Fp- (up) or mock- inoculation (down). DEGs were determined with the threshold of FDR ≤ 0.05 and |log2 fold-change|≥ 1 or 'inf' (one of the comparative objects did not express and the other did)

4.3.3 SNPs between the 'R' and 'S' isolines across the three 1H_NIL pairs

In total, 2,753 non-redundant homozygous SNPs were detected between the 'R' and 'S' isolines. The number of SNPs detected from 1H_NIL2 was more than twice compared with those detected from either of the other two NIL pairs. Of these SNPs, 293 were common among the three pairs of the 1H_NILs. As expected, the majority of the SNPs shared among the three NIL pairs located at the distal end of chromosome arm 1HL where *Qcrs.cpi-1H* resides (Fig. 4.4). They spanned a physical distance of ~ 11.0 Mbp (Fig. 4.5a).

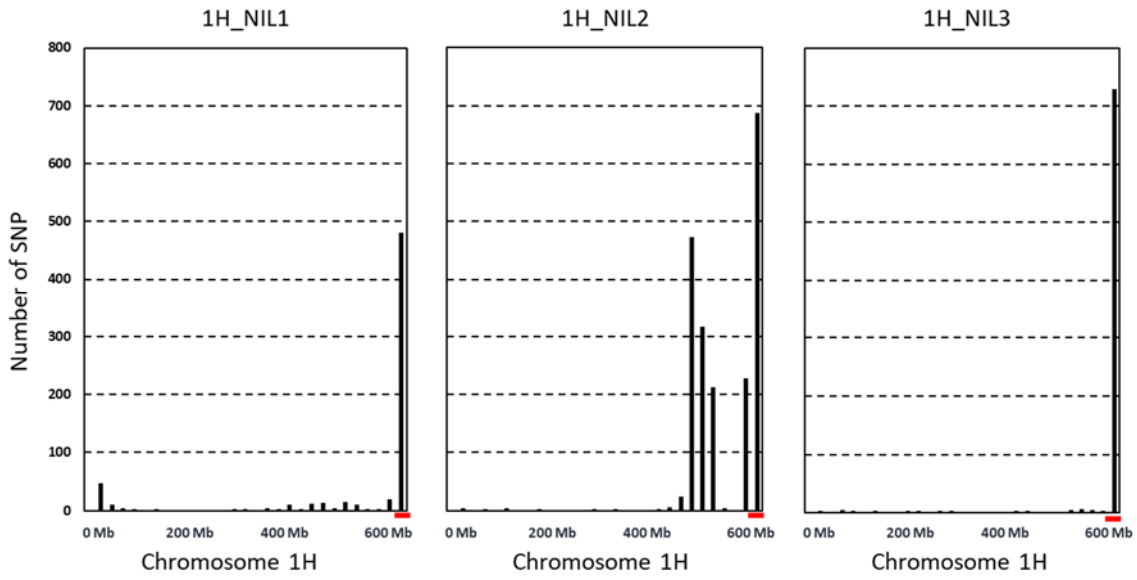


Fig. 4.4 Distribution of SNPs in the expressed genes along chromosome 1H in three pairs of the 1H_NILs. Vertical axis shows number of SNPs. Horizontal axis shows chromosome 1H from short (left) to long (right) arm in base pairs (bp). Red bars represent the candidate region harbouring the FCR resistant locus Qcrs.cpi-1H.

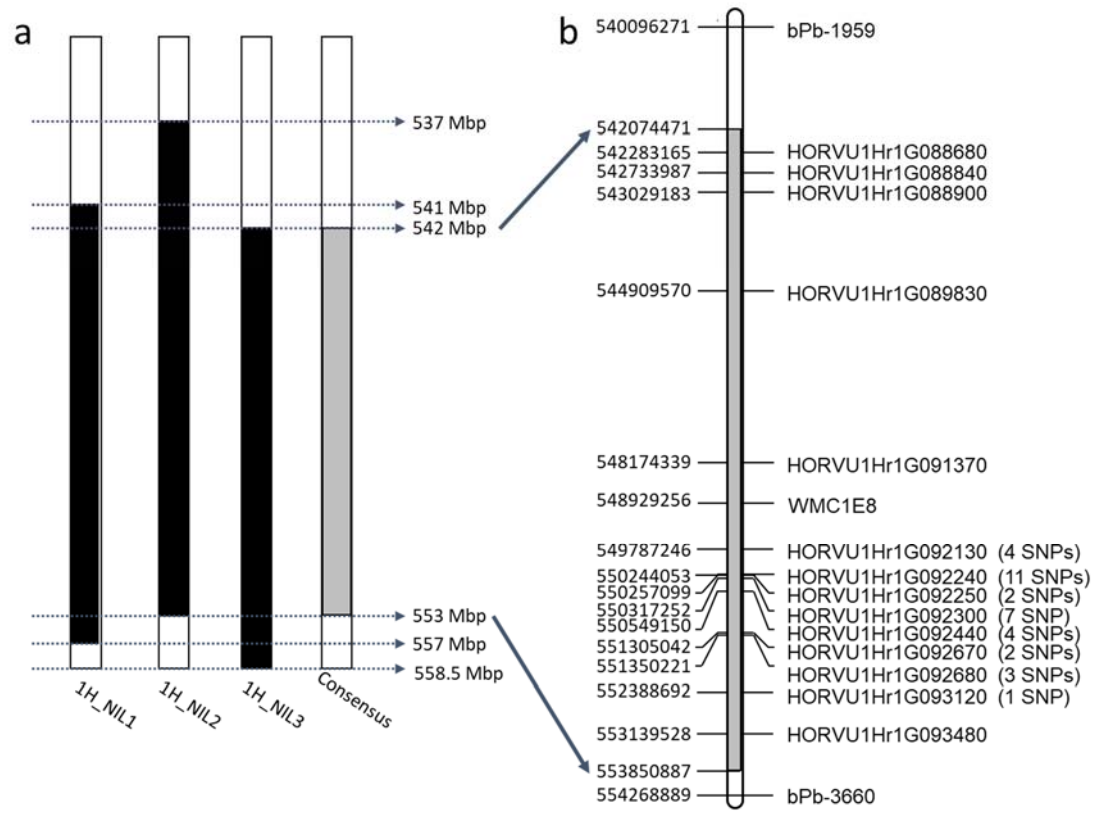


Fig. 4.5 Physical distribution of DEGs within the consensus SNP-enriched region. **a** The physical range of SNP-enriched regions. Black boxes indicate the regions defined by SNPs within each 1H_NIL pair; the grey box represents the consensus region. **b** Physical distribution of DEGs shared among the three comparisons within the consensus region. The initial QTL region was flanked by bPb-1595 and bPb-3660. SNP-up/down indicate the borders of the consensus region. The numbers of SNPs identified within genes were in brackets.

4.3.4 DEGs with SNPs between the resistant and susceptible isolines targeting the *Qcrs.cpi-1H* locus

Based on the reference genome of barley cv. Morex, 266 HC genes were identified within the common interval across three 1H_NIL pairs. Among these HC genes, 56 contained SNPs and 14 were differentially expressed between the isolines for at least one of the NIL pairs (Fig. 4.5b; Table S4.3). Notably, five protein-coding genes were not only differentially expressed across the three NIL pairs but also carried SNPs led to non-synonymous variations (Table 4.3 and Table S4.4). These protein-coding genes should form the primary targets in identifying candidate genes underlying FCR resistance at this locus.

Table 4.3 Expression patterns of five DEGs bearing non-synonymous SNPs located in the interval harbouring the FCR resistant locus *Qcrs.cpi-1H*

Gene ID	Gene Description [a]	Number of Non-synonymous SNPs	Pattern of differential expression
HORVU1Hr1G092130	WRKYDNA-binding protein 23	1	Upregulated in 3 S isolines post inoculation
HORVU1Hr1G092240	Glucan endo-1,3-beta-glucosidase13	4	Upregulated in 3 R isolines post inoculation
HORVU1Hr1G092250	Receptor-like kinase	1	Upregulated in 3 R and 3 S isolines post inoculation
HORVU1Hr1G092300	Receptor-like kinase	6	Upregulated in 3 R post inoculation
HORVU1Hr1G092440	P-loop containing nucleoside triphosphate hydrolases super family protein	4	Upregulated in 3 S isolines post inoculation

[a] Gene descriptions were retrieved from the annotation file of the genome of barley cv. Morex

4.4 Discussion

FCR is a chronic disease for cereal production in semi-arid regions worldwide. It has long been recognised that breeding and growing resistant varieties have to form an integral part in the effect of effectively reducing damages from the disease. Previous studies also show that strong interactions between FCR severity and several characteristics including flowering time and plant height exist thus QTL detected from mapping populations need to be validated. In the study reported here, we successfully validated the QTL on chromosome arm 1HL by developing and assessing NILs targeting the locus. DEGs with SNPs shared by three pairs of the NILs further delineated the locus to an interval of about 11.0 Mbp. They would be invaluable for fine mapping the locus and cloning the gene(s) underlying its resistance. SNPs in several of the DEGs lead to amino acid changes and they would be primary targets in investigating the mechanism of FCR resistance.

It is of note that significant variation was found in the numbers of DEGs detected among the three pairs of NILs assessed. Previous studies showed that FCR development can be affected by various characteristics including plant height (Bai and Liu 2015; Chen et al. 2014; Li et al. 2009; Liu et al. 2010) and flowering time (Chen et al. 2013a; Liu et al. 2012a; Liu et al. 2012b). Each of the NIL pairs used in this study was developed from a different heterozygous plant based on the profile of a single marker. This method ensured that different NIL pairs, including those from the same population, would have different genetic backgrounds. The different genetic backgrounds would lead to difference in FCR development at any given time point. In other words, although symptom of FCR infection was not visually observable for any of the NILs at 4 dpi when the samples for RNA-seq were taken, the advancement of FCR development among them must be different.

The interactions between FCR severity and other characteristics may also contributed to the difference in the effects of the 1HL locus between the use of NILs as described in this study and that based on QTL mapping (Chen et al. 2013b). In addition to the targeted trait, many other characteristics likely also segregate in populations routinely used for QTL mapping. They include populations of recombinant inbred lines and doubled haploid lines. In essence, a targeted locus is always assessed in different genetic backgrounds in QTL mapping studies, making its accurate assessment difficult. In the contrary, the two isolines forming each NIL pair differ mainly by the targeted locus. The fact that assessments for any characteristics can

be carried out by comparing two isolines only must also contribute to the likelihood that more accurate assessment can be achieved by using NILs.

It is also of note that significant difference in FCR resistance was not detected between isolines for three of the eight pairs of putative NILs developed in this study. Different from the method of using markers flanking the targeted locus (Pumphrey et al. 2007a), we used only one linked marker obtained from a QTL mapping study [10] in developing the NILs. As discussed in earlier reports (Habib et al. 2016; Ma et al. 2012), the approach of using a single linked marker is preferred as it should reduce the sizes of chromosomal segments differentiating the isolines for NILs obtained. However, QTL mapping studies have only limited resolution (Paterson et al. 1988) thus markers obtained from such studies may not be tightly linked with a targeted locus. Clearly, recombination between the linked marker and its target may occur, resulting in false NILs.

Within the targeted interval of the 1HL locus, five protein-coding genes are highly interesting due not only to their patterns of expression among the NILs but also the fact that they contain nonsynonymous SNPs. They are known to be involved in plant-pathogen interaction or abiotic stress (i.e. drought) which facilitates *F. pseudograminearum* infection. They include the two receptor-like kinase (RLK) genes (*HORVU1Hr1G092250* and *HORVU1Hr1G092300*) which are involved in the immune systems in various plant species (Marone et al. 2013). RLK locates on either the plasma or cytoplasmic membrane and are responsible for recognizing elicitor, usually small secreted protein, generated by pathogens. The perception of elicitor often triggers a fierce hypersensitive response which can cause programmed cell death (Krattinger and Keller 2016). Another one is the gene for glucan endo-1,3,-beta-glucosidase (*HORVU1Hr1G092240*) which plays an important role in defence against pathogen infection (Beffa et al. 1993). Its expression has been detected in the response to biotic stress in various plant species (Faghani et al. 2015; Su et al. 2016). *HORVU1Hr1G092440* encoding a P-loop containing nucleoside triphosphate hydrolases (P-loop NTPase) protein is also among the DEGs with SNPs located in the targeted interval. Previous results showed that this gene negatively regulates plant defence response in both rice and *Arabidopsis* (Cheung et al. 2010; Cheung et al. 2013). Once bonded with ATP, *OsYchF1*, a P-loop NTPase in rice, contributes to resistance to biotic stress (Cheung et al. 2016).

It is also interesting to note that one of the DEGs with SNPs located in the targeted interval confers tolerance to drought. This is *HORVU1Hr1G092130* which codes a WRKY

transcription factor which plays a key role in signaling in the defense response to biotic and abiotic stress (Birkenbihl et al. 2017; Eulgem et al. 2000). A homolog of *HORVU1Hr1G092130* in rice, *Os05g0583000* was strongly induced during drought response (Shin et al. 2016). Over-expression of *Os05g0583000* coding sequence in *Arabidopsis* provided improved drought tolerance (Song et al. 2009). The presence of this gene related to drought tolerance is not a surprise as the relationship between drought stress and *Fusarium* crown rot severity in agricultural systems has been well documented. FCR causes severe yield loss mainly in semi-arid regions (Chakraborty et al. 2006) and drought stress forms part of the procedures in FCR assay, which was also performed in the current study, in both wheat (Ma et al. 2010; Zheng et al. 2014) and barley (Chen et al. 2013a; Chen et al. 2013b; Gao et al. 2019b). As such, it is not unexpected that the causal gene of *Qcrs.cpi-1H* may decrease FCR disease expression through improved drought stress tolerance rather than classical disease resistance mechanisms.

Based on the DEGs detected in this work, it also seems unlikely that the mechanism for resistance is driven by differences in classical resistance mechanisms previously described as important for defence against *Fusarium* pathogens (Table S5). The *Fusarium* mycotoxin, deoxynivalenol has been shown to be required for full virulence of *F. pseudograminearum* when infecting wheat and *Brachypodium* (Powell et al. 2017a; Powell et al. 2017b). Detoxification of deoxynivalenol has been strongly implicated in defence against *F. graminearum* causing *Fusarium* head blight with DON detoxifying UDP glycosyltransferases identified in wheat, barley and *Brachypodium* (Gatti et al. 2018; Schweiger et al. 2010; Schweiger et al. 2013). The UDP-glycosyltransferase detoxifying DON in barley (*HORVU5Hr1G047150*) (Li et al. 2017; Schweiger et al. 2010) was not found to be differentially expressed between or showing SNPs differences between R or S isolines in the current study (Table S5). Previous studies have also shown that induced systemic resistance mechanisms are involved in response to *F. pseudograminearum* infection (Desmond et al. 2008). Key markers for systemic acquired or induced systemic resistance, such as genes encoding jasmonate biosynthetic enzymes, salicylic acid biosynthetic enzymes and pathogenesis related proteins, were differentially expressed in response to infection across both resistant and susceptible isolines to similar magnitudes (Table S5). Therefore, from comparison of molecular responses observed in resistant and susceptible isolines, we did not find any inference that the effect of the 1HL locus occurs through previously characterised quantitative

resistance mechanisms. We thus conclude that resistance mediated by the 1HL resistance locus may provide a highly novel FCR resistance source in barley.

4.5 Conclusions

In this study, we developed five pairs of NILs targeting the FCR resistance locus *Qcrs.cpi-1H*. Phenotyping these NIL found that the resistant allele at *Qcrs.cpi-1H* could significantly reduce FCR severity. Gene expression and SNP analysis of transcriptomic data derived from three pairs of the 1H_NILs delineated the *Qcrs.cpi-1H* locus into an about 11 Mbp interval containing 56 genes with SNP(s). Of these genes, five DEGs bearing non-synonymous SNPs form primary targets in identifying gene(s) underlying the *Qcrs.cpi-1H* locus.

Chapter 5 Delineating a locus conferring *Fusarium* crown rot resistance on chromosome arm 1HL in barley by developing and analysing a large population derived from near isogenic lines[3]

[3] This paper has been published as: Gao, S., Zheng, Z., Hu, H., Jiang, Y., Liu, M., Stiller, J., Zhou, M., Liu, C., 2020. Delineating a locus conferring Fusarium crown rot resistance on chromosome arm 1HL in barley by developing and analysing a large population derived from near isogenic lines. The Crop Journal. https://doi.org/10.1016/j.cj.2020.03.008

5.1 Introduction

Fusarium crown rot (FCR), caused by various species of *Fusarium*, is a chronic disease for cereal production in arid and semi-arid cropping regions worldwide (Chakraborty et al. 2006; Hogg et al. 2010). Initial infection of FCR is characterized by brown lesions in the crown and lower stem regions and inside leaf sheaths. Tillering ability of plants is affected. Under moisture stress, especially during the period between anthesis to milky ripening, 'whiteheads' containing shriveled or no grains could occur in FCR infected fields. As a result, grain yield and hence crop value can be significantly affected (Chakraborty et al. 2006; Murray and Brennan 2010). Significant yield losses due to this disease on wheat and barley has been reported in numerous countries (Chakraborty et al. 2010; Hameed et al. 2012; Li et al. 2012; Xu et al. 2017). Incidence and severity of FCR have been exacerbated in Australia in recent years, likely due to the increase in the intensity of cereal production for economic reasons and the wide adoption of reduced tillage for moisture conservation (Chakraborty et al. 2006). Reduced frequency of precipitation and increased temperature during crop growth have also been identified as possible factors contributing to the exacerbation of the disease (Singh et al. 2009).

Reducing inoculum load, including crop rotation and stubble burning, has been the focus of management practices for reducing FCR damage (Cook 2001; Kirkegaard et al. 2004). This is based on the belief that physical contact of the stem base with infested stubble of the proceeding years facilitates the disease infection (Burgess 2005; Burgess 2014). These practices, however, have serious limitations. Stubble burning is not only a serious environmental concern but also leads to loss of soil moisture and wiping out beneficial soil microbes. Crop rotation is not always practical as the FCR pathogens can survive several years in stubble (Burgess 2005; Burgess 2014) and growing less valuable crops may also lead to the loss of income. With the

wide adoption of precision farming, inter-row sowing has also been recommended for minimizing yield loss from the disease (Simpfendorfer et al. 2012).

It was realized for a long time that growing resistant varieties is a critical component in effectively managing FCR (Purss 1966). Sources of resistance were identified and numerous QTL conferring FCR resistance have been detected in both wheat and barley (Liu and Ogbonnaya 2015). QTL mapping, however, can only provide limited resolution (Paterson et al. 1988) thus markers obtained from such studies cannot be reliably used for marker-assisted selection. This is because of that, together with those for the targeted trait, many other loci for a wide range of characteristics may also segregate in a given population used for QTL mapping. Plant height and growth rate have been found affecting FCR assessment in both barley (Bai and Liu 2015; Chen et al. 2013a; Chen et al. 2013b; Chen et al. 2014; Li et al. 2009; Liu et al. 2010) and wheat (Ma et al. 2010; Zheng et al. 2014). Clearly, segregations of these characteristics in a mapping population would make accurate assessment of FCR severity difficult.

As the two isolines for a given pair of near-isogenic lines (NILs) share a similar genetic background which is essentially fixed, the difference between them is mainly due to the differences in the targeted locus. The unique features of NILs make them highly effective in validating QTL conferring various characteristics (Pumphrey et al. 2007b). Combined with techniques that can speed up life cycles (Liu et al. 2016a; Yan et al. 2017; Yao et al. 2016; Zheng et al. 2013), NILs can now be conveniently and quickly obtained for different crop species. These techniques have been used to develop NILs targeting loci conferring FCR resistance in both wheat (Ma et al. 2012) and barley (Habib et al. 2016). Importantly, a NIL-derived population can be conveniently used to develop markers tightly linked with a given locus as, different from those routinely used for QTL mapping, such populations segregate mainly for the targeted locus under investigation. These approaches have been used to investigate an FCR resistance locus on chromosome arm 1HL. This locus, termed as *Qcrs.cpi-1H*, was initially identified from a landrace originated from Japan. It explained up to 33.4% of FCR severity variance (Chen et al. 2013b). NILs targeting this QTL were generated and used to validate its effects in different genetic backgrounds, and transcriptomic differences between the resistant and susceptible lines for three pairs of the NILs were investigated in a previous study (Gao et al. 2019c). In the study reported here, we delineated the locus in a refined interval and obtained markers co-segregating with the locus by generating and characterising a large

NIL-derived population. By analysing differentially expressed genes located in the refined interval, a small number of candidate genes underlying FCR resistance at this locus were also identified.

5.2 Materials and method

5.2.1 Plant materials

A NIL-derived population consisting of 1,180 lines was generated and used in this study. This population was derived from five different heterozygous F7 plants obtained in generating the NILs (1H_NIL1) targeting this locus based on the marker *WMC1E8* (Gao et al. 2019c). A single-seed-descent approach was used to process the F7 heterozygous plants by five further rounds of self-pollination using the fast-generation method (Zheng et al. 2013). Seeds from each of the lines were then increased in large pots and used for this study. The population was processed in glasshouses at Queensland Bioscience Precinct in St Lucia, Australia.

5.2.2 Preparation of inoculum and evaluation of FCR resistance

A highly aggressive isolate of *F. pseudograminearum,* CS3096, was used to assess FCR resistance in this study. Methods used for inoculum preparation, FCR inoculation and disease severity assessment were conducted following the methods described by Li et al. (2008). Briefly, inoculum was prepared in plates of ½ strength potato dextrose agar. Inoculated plates were kept for 12 days at room temperature before the mycelium were scraped and discarded. The plates were then incubated for a further 7-12 days under a combination of cool white and black fluorescent lights with 12-h photoperiod. Spores were then harvested, and the concentration of spore suspension was adjusted to 1×10^6 spores/ml. The spore suspension was stored in a -20 °C freezer. Tween 20 was added (0.1%v/v) to the spore suspension prior to use.

Seeds were germinated in Petri dishes on three layers of filter paper saturated with water. Seedlings of 3-day-old were immersed in the spore suspension for 1 min. Two seedlings were planted into a 3 cm square punnet (Rite Grow Kwik Pots, Garden City Plastics, Australia) containing sterilized University of California mix C (50% sand and 50 % peat v/v). The punnets were arranged in a randomized block design and placed in a controlled environment facility (CEF). Settings for the CEF were: 25/18 (±1) °C day/night temperature, 65/80% (±5)% day/night relative humidity, and a 14-h photoperiod with 500 µmol $m^{-2}S^{-1}$ photon flux density

at the level of the plant canopy. Inoculated seedlings were watered only when they started to wilt.

To confirm the location of the targeted locus and identify markers flanking it, a subpopulation containing 88 of the NIL-derived lines was assessed in three independent inoculation trials. Each trial contained two replicates and 14 seedlings were used in each of the replicates. Markers flanking the targeted locus developed based on the subpopulation were then used to identify recombinant lines from the whole NIL-derived population. Five independent trials were conducted on the recombinant lines identified. The resistant and susceptible isolines of the NIL pair 1H_NIL1 were used as positive and negative controls, respectively, in each of the inoculation trials. FCR severity was assessed four weeks post inoculation with a 0 – 5 scorings, where "0" stands for no symptom and "5" for whole seedling completely necrotic.

5.2.3 Phenotypic data analysis

Statistical analyses of all phenotypic data were performed using R (R core team 2013). For each trial, the following mixed-effect model was used: $Yij = \mu + ri + gj + wij$. Where: Yij = trait value on the jth genotype in the ith replication; μ = general mean; ri = effect due to ith replication; gj = effect due to the jth genotype; wij = error or genotype by replication interaction, where genotype was treated as a fixed effect and that of replicate as random. The disease scores from all seedlings for each of the NIL-derived lines were averaged to determine whether a given line is resistant (<2.5) or susceptible (>2.5) to FCR infection.

5.2.4 Identification of the targeted interval and marker development

The *Qcrs.cpi-1H* had been mapped into a physical interval of ~11 Mbp in a previous study based on RNA-seq analysis against several sets of the NILs targeting the locus (Gao et al. 2019c). Insertion/deletion (indel) and kompetitive allele specific PCR (KASP) markers targeting this interval were developed and used in this study. Indel markers were developed based on the variants between the pseudomolecule of 'Morex' (Mascher et al. 2017) and an assembly of a wild barley (*Hordeum spontaneum*) genotype *AWCS276* (Liu M. et al., unpublished) using SSR-finder (https://github.com/GouXiangJian/SSR_finder). For KASP markers, SNPs within the interval were detected using RNA-seq sequence from three of these NIL pairs (Gao et al. 2019c) on CLC genomic workbench platform V11.0 (CLC Bio, Aarhus, Denmark) and primers were designed based on results from the primer-blast (Ye et al. 2012).

All the primers and their sequences are listed in Tables S1 and S2. MSTmap Online (Wu et al. 2008) was used to build linkage maps with Kosambi function. The genetic linkage map was plotted using MapDrawJZ (https://github.com/pinbo/MapDrawJZ), a modified version of MapDraw V2.1 (Liu and Meng 2003).

5.2.5 DNA extraction and genotyping

Leaf tissue from each line of the NIL-derived population was collected and vacuum dried for DNA extraction using the CTAB protocol (Porebski et al. 1997). KASP assay were conducted using 384-well set on the Vii^{TM} 7 Real-Time PCR system (Applied Biosystems) following the "KASP genotyping trial kit user guide" (https://biosearch-cdn.azureedge.net/assetsv6/KASP-genotyping-trial-manual.pdf) and "Guide to running KASP genotyping reactions on the ABI Viia7 instrument" (https://biosearch-cdn.azureedge.net/assetsv6/running-KASP-on-ABI-Viia7.pdf). Indel makers were assessed according to the method described by Zheng et al (2015).

5.2.6 Identification of candidate genes, nonsynonymous SNPs and collinearity analysis

Annotations of both high confidence (HC) and low confidence (LC) genes in the genomic interval defined by the two flanking markers for the *Qcrs.cpi-1H* locus were extracted from the barley genome explore (BARLEX) (https://barlex.barleysequence.org). SNPs contained in these genes were identified using snpEff 4.3q (Cingolani et al. 2012). The variant database was built based on the international barley reference genome of 'Morex' and its annotation file (Mascher et al. 2017). Orthologs for candidate genes surrounding the 1HL locus in *Brachypodium distachyon* and rice (*Oryza Sativa* L.*)* were extracted using Ensembl Plant BioMart (Kinsella et al. 2011).

5.3 Result

5.3.1 Validation of the chromosomal interval containing *Qcrs.cpi-1H* based on analysing the subpopulation with 88 NIL-derived lines

Based on results from the RNA-seq analysis (Gao et al. 2019c), four indel markers targeting the interval were developed based on sequence differences between the resistant and

susceptible NILs. Together with *WMC1E8* (the marker initially used for generating the NILs), the five markers all segregated in the subpopulation (Fig. 5.1a). Linkage analysis showed that they spanned a genetic distance of ~7.3 cM and covered a length of ~5.2 Mbp based on the barley reference genome of 'Morex'. FCR severity assessment of this subpopulation showed that all lines fell into a binary pattern, i.e. their FCR severity scores belonged to either the resistant or susceptible classes (Fig. S5.1). Combined the marker profiles and the phenotypic data mapped *Qcrs.cpi-1H* into a ~2.6 Mbp interval flanked by *WMC1E8* and *Sgs_5514* (Fig. 5.1a).

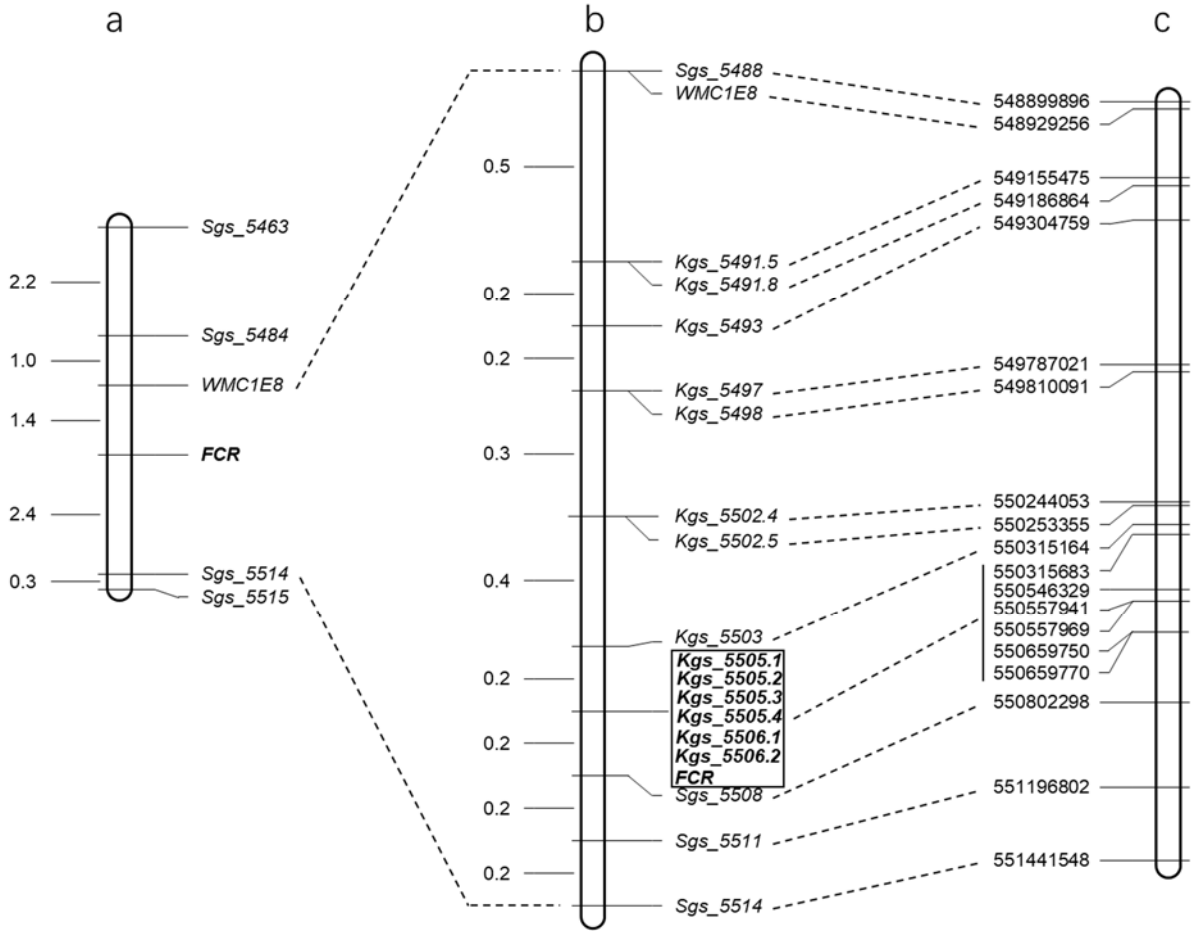


Fig. 5.1 Genetic and physical maps surrounding the Fusarium crown rot resistance locus *Qcrs.cpi-1H* in barley. (a) The targeted interval based on the assessment of a subpopulation consisting of 88 NIL-derived lines. (b) The high-density linkage map surrounding *Qcrs.cpi-1H* based on the analysis of the whole population consisting of 1,182 NIL-derived lines. Markers co-segregating with the locus are in bold and placed in a box. (c) Physical positions of markers surrounding *Qcrs.cpi-1H* on the 1H pseudomolecule of the 'Morex' genome.

5.3.2 Fine mapping of the *Qcrs.cpi-1H* locus using the NIL-derived population consisting of 1,180 lines

WMC1E8 and *Sgs_5514* were used to screen the whole NIL-derived population containing of 1,180-lines. Twenty-five recombinant lines were identified between these two markers. FCR assessments against these recombinants found that six of them were FCR resistant and the other 19 susceptible (Fig. 5.2). The difference in FCR severity between the two groups of recombinants was highly significant ($P < 0.01$; student's *t*-test).

To construct a high-density map spanning the targeted interval, 14 KASP markers and three indel markers between the two flanking markers were generated and assessed against the 25 recombinant lines. Linkage analysis showed that the FCR locus co-segregated with six of these markers and it was placed at 0.2 cM proximal to *Kgs_5503* and 0.2 cM distal to *Sgs_5508* (Figs. 5.1b, 5.2). The linkage order of the markers surrounding *Qcrs.cpi-1H* were identical with their relative physical positions in the 'Morex' genome (Fig 5.1b, c). Based on the physical positions of its flanking markers, *Kgs_5503* and *Sgs_5508*, *Qcrs.cpi-1H* was delimited to a ~487 Kbp genome interval from 550.3 Mbp to 550.8 Mbp on the 1H pseudomolecule of 'Morex'.

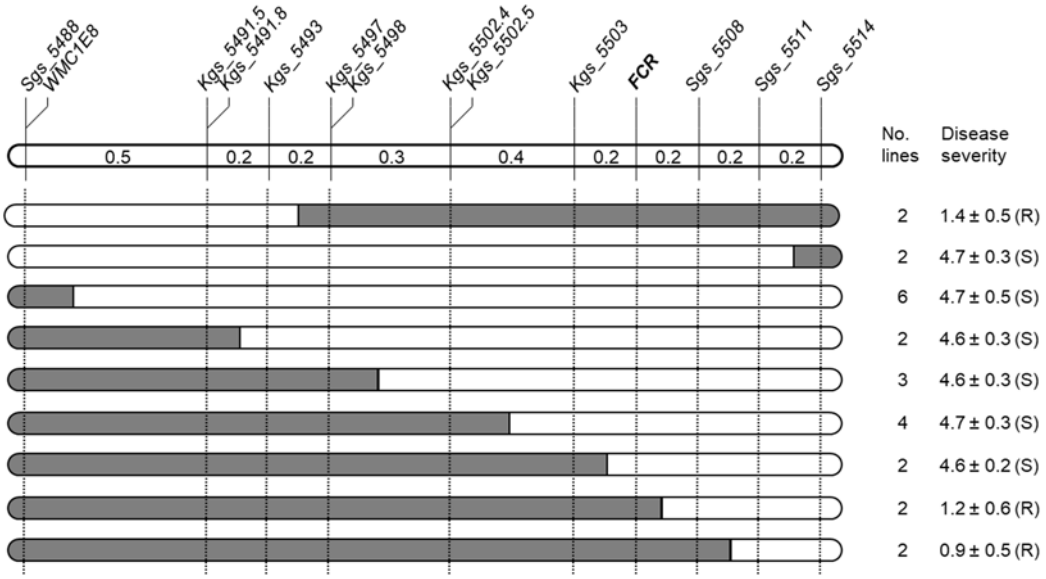


Fig. 5.2 Genotypes and phenotypes of the key recombinant lines identified with markers surrounding the *Qcrs.cpi-1H* locus. The solid regions represent alleles from the resistant parent and the empty ones for alleles from the susceptible parent. Numbers of recombinant lines and corresponding FCR severity (mean ± SE) were provided on the right-hand side of the diagram. The numbers between makers stand for genetic distance (cM).

5.3.3 Identification of candidate genes in the targeted region

Based on the 'Morex' genome, the targeted interval contained 13 high confidence (HC) and 15 low confidence (LC) genes (Table S5.3). The LC genes were not taken into further consideration due to lack of clear functional annotation. Expression profiles and single nucleotide variants of the 13 HC gene were examined using RNA-seq data generated from the three pairs of the NILs targeting *Qcrs.cpi-1H* (Gao et al. 2019c). Six of these HC genes expressed in at least one of the NIL pairs used (Fig. 5.3). SNPs between resistant and susceptible isolines were identified in five of the HC genes. SNPs in three of the genes led to non-synonymous variations (Fig. 5.3; Table S5.4). Two of these genes *HORVU1Hr1G092310* and *HORVU1Hr1G092440*, carrying non-synonymous SNPs were up-regulated in each of the NILs following FCR inoculation. The third gene containing a non-synonymous SNP, *HORVU1Hr1G092550,* encodes a receptor-like kinase. This gene expressed consistently in all the NILs, with or without FCR inoculation (Fig. S5.3).

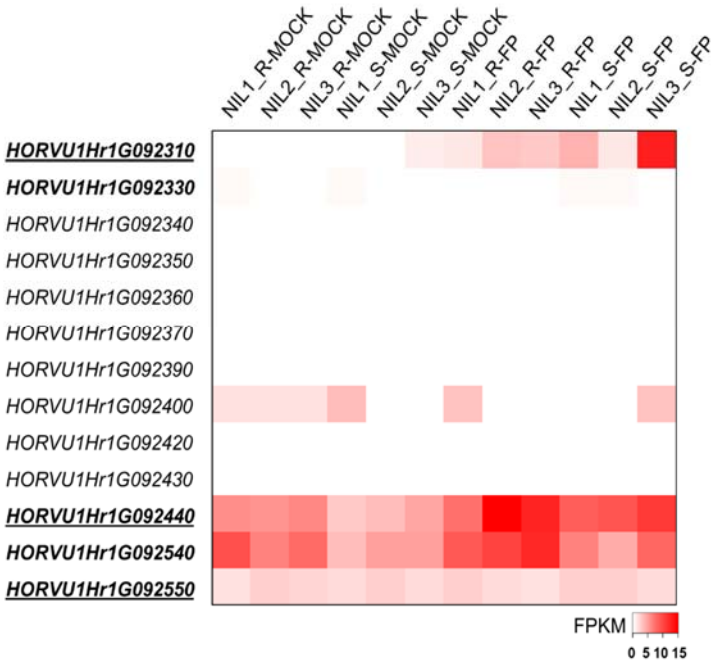


Fig. 5.3 Expression profiles of the high confidence (HC) genes among the three pairs of near-isogenic lines targeting the FCR resistance locus *Qcrs.cpi-1H*. The heatmap shows the levels of absolute expression of these genes in resistant (R) and susceptible (S) NILs 4 days post water (MOCK) or *Fusarium pseudograniearum* (FP) treatment. Genes carrying SNP(s) were in bold and the genes containing non-synonymous SNP(s) were underlined.

5.3.4 Collinearity between the genes in the targeted interval and those in *Brachypodium* and rice

Collinearity for genes located in the identified interval containing the targeted locus *Qcrs.cpi-1H* was assessed with their corresponding regions in the genomes of *Brachypodium* and rice (Table 5.1; Fig. S5.2). This analysis found that gene collinearity in this interval among these three species was poor. Orthologs for only three of the 13 barley genes were found in the corresponding interval in the *Brachypodium* genome, and the orders of the genes were different between the two genomes. Of the 10 genes in the *Brachypodium* genome, orthologs for seven of them were not found in the barley genome. Similarly, orthologs for only four of the 13 barley genes in the targeted interval were detected in the corresponding region of the rice genome and the orders were different between the two genomes. Orthologs for six of the 10 genes in the rice genome were not found in the barley genome (Table 1, Fig. S2). Of the *Brachypodium* and rice genes absent in the barley genome, six were annotated with a wide range of functions. Two of them were found in both genomes of *Brachypodium* and rice, one encoding a late embryogenesis abundant protein and the other with unknown functions (Table 5.1).

Table 5.1 High confidence genes surrounding the *Qcrs.cpi-1H* locus and their orthologs in the corresponding genomic regions in *Brachypodium distachyon* and *Oryza sativa*

Brachypodium distachyon [a]	*Hordeum vulgare* [a]	*Oryza sativa* [a]	Putative function [b]
BRADI_2g15400v3	-	-	F-box family protein
-	-	*Os05g0583551*	Hypothetical conserved gene
-	*HORVU1Hr1G092310*	-	Glucan endo-1,3-beta-glucosidase 13
-	*HORVU1Hr1G092330*	-	Sugar transporter protein 7
-	*HORVU1Hr1G092340*	*Os05g0584900*	S-type anion channel SLAH2
-	*HORVU1Hr1G092350*	-	Sulfate transporter 4;2
BRADI_2g15405v3	*HORVU1Hr1G092360*	*Os05g0583600*	WRKY DNA-binding protein 27
-	-	*Os05g0583950*	Hypothetical conserved gene
BRADI_2g15410v3	-	*Os05g0584200*	Late embryogenesis abundant protein
-	*HORVU1Hr1G092370*	-	Ycf68
-	*HORVU1Hr1G092390*	-	Cell wall-associated hydrolase
-	*HORVU1Hr1G092400*	-	Cell wall-associated hydrolase
-	*HORVU1Hr1G092420*	-	30S ribosomal protein S15, chloroplastic
-	*HORVU1Hr1G092430*	-	Unknown function
-	-	*Os05g0584750*	Hypothetical protein
BRADI_2g15490v3	*HORVU1Hr1G092440*	*Os05g0584600*	P-loop containing nucleoside triphosphate hydrolases superfamily protein
BRADI_2g15480v3	-	-	Similar to Dihydrodipicolinate synthase 1, chloroplast precursor
-	-	*Os05g0584450*	Hypothetical gene
BRADI_2g15471v3	-	*Os05g0584400*	Conserved hypothetical protein
BRADI_2g15460v3	-	-	Glycosyl transferase
BRADI_2g15450v3	-	-	Glycosyl transferase
BRADI_2g15440v3	-	-	Glycosyl transferase

BRADI_2g15420v3	*HORVU1Hr1G092540*	*Os05g0584300*	Late embryogenesis abundant hydroxyproline-rich glycoprotein family
-	*HORVU1Hr1G092550*	-	LRR receptor-like serine/threonine-protein kinase GSO2

[a] The high confidence genes in barley were based on their physical order on the 1H pseudomolecule of 'Morex'; positions of the *Brachypodium* and rice orthologs were adjusted according to the positions of their orthologs in the barley genome;

[b] Putative functions of the genes were retrieved from Ensembl Plants (http://plants.ensembl.org/index.html).

5.4 Discussion

FCR causes severe damage to cereal production in semi-arid areas worldwide and growing resistant varieties is recognised as an essential component to effectively managing the disease. By developing and assessing a NIL-derived population consisting of 1,180 lines, we delineated a major FCR locus *Qcrs.cpi-1H* into a genomic interval of about 487 kb and developed six co-segregating markers. Based on the 'Morex' genome, 13 HC genes were contained in the interval. SNPs in three of the genes resulted in non-synonymous variants in proteins they encode. Two of the genes were up-regulated following FCR inoculation and the other was consistently expressed with or without FCR infection in both the resistant and susceptible isolines assessed. These results should be invaluable not only in incorporating the resistance locus into breeding programs but also in identifying the causal gene(s) underlying the locus.

Reproducible and reliable phenotypic data are critical for high quality mapping of any locus (Cuthbert et al. 2006). Previous studies have repeatedly shown that several characteristics affect the accurate assessment of FCR severity. This includes both plant height (Li et al. 2010; Liu et al. 2010; Ma et al. 2010) and flowering time (Chen et al. 2013a; Zheng et al. 2015). For minimizing the interference from the segregations of these non-targeted characteristics in FCR assessment, a NIL-derived population targeting the *Qcrs.cpi-1H* locus was developed and used in the study reported here. As expected, FCR severities among the lines of this large NIL-derived population were easily placed into ether a resistant or susceptible class, making it possible to accurately place the targeted locus in a well-defined genomic interval.

Of the 13 HC genes located within the targeted interval containing the FCR resistance locus *Qcrs.cpi-1H,* six were detected in the transcriptomic data obtained in studying genes responsive to FCR infection using three pairs of the NILs targeting this locus (Gao et al. 2019c). Compairing the resistant and susceptible NILs, SNPs were identified leading to non-synonymous variations in three HC genes and these genes play critical roles in plant defence to pathogens. One of the genes, *HORVU1Hr1G092550*, encodes a receptor-like kinase (RLK) which has been identified in various immune systems of plants (Marone et al. 2013). RLK, usually located on either plasma or cytoplasmic membrane, is able to recognize elicitors generated by pathogens and triggers downstream defence responses in the plant to resist pathogens (Krattinger and Keller 2016). The second gene, *HORVU1Hr1G092310*, encodes a glucan endo-1,3-beta-glucosidase which is known to be involved in systemic acquired

resistance (Ryals et al. 1996). This enzyme plays an important role in seed plant defence against pathogen attack through the degradation of fungal cell wall polysaccharides (Høj et al. 1989). The third gene *HORVU1Hr1G092440* encodes a P-loop containing nucleoside triphosphate hydrolase (P-loop NTPase) which is known to negatively regulate the abiotic stress and plant defence response in both rice and *Arabidopsis* (Cheung et al. 2010; Cheung et al. 2013).

Clearly, the three genes with non-synonymous variations between the resistant and susceptible NILs must be carefully examined in identifying gene(s) underlying FCR resistance at the targeted locus. However, recent studies show that non-classical NBS-LRR genes can also be responsible for resistance to a wide range of pathogens in plants (Milne et al. 2019; Moore et al. 2015). They include the *Fhb1* gene conferring *Fusarium* head blight (FHB) in wheat (Li et al. 2019; Su et al. 2019a). Results from previous studies showed that *Fusarium* pathogens causing FHB can also lead to FCR (Chakraborty et al. 2006). Similar to the situation for FHB (Van Eeuwijk et al. 1995), host resistance to FCR is also not pathogen species-specific (Li et al. 2010; Ma et al. 2010). The common aetiology between FHB and FCR raises the possibility that resistance to the latter may also be conferred by non-classical NBS-LRR genes.

It has become clear in recent studies that large numbers of genes in a given species are 'dispensable', thus gene(s) underlying the FCR resistance locus *Qcrs.cpi-1H* may not necessarily be present in the genome of 'Morex' which is highly susceptible to FCR. For examples, the components of dispensable genes are about 50% in maize (*Zea mays* L.) (Hirsch et al. 2014), 43% in rice (*Oryza sativa* L.) (Sun et al. 2016); 36% in bread wheat (*Triticum aestivum* L.) (Liu et al. 2016b; Montenegro et al. 2017) and 38% in barley (Ma et al. 2019). To identify additional genes likely located in the targeted interval, we analysed the corresponding genome regions in both *Brachypodium* and rice. This analysis found that synteny for the targeted genomic regions among the three species is poor. Several of the genes in the 'Morex' genome were not found in the orthologous regions of either *Brachypodium* or rice, and the orders for the few shared orthologs are often different. Nevertheless, the ten genes present in either *Brachypodium* or rice but absent in the 'Morex' genome should also be considered in identifying genes underlying the FCR resistance locus.

Chapter 6 Identifying barley pan-genome sequence anchors using genetic mapping and machine learning[4]

[4] This paper has been published as: Gao, S., Wu, J., Stiller, J., Zheng, Z., Zhou, M., Wang, Y., and Liu, C., 2020. Identifying barley pan-genome sequence anchors using genetic mapping and machine learning. Theoretical and Applied Genetics, https://doi.org/10.1007/s00122-020-03615-y

6.1 Introduction

With the progress in genome sequencing, it has become clear that an individual genotype contains only a proportion of the genes found in a given species. The phenomenon was initially noticed in bacterial genomes two decades ago. To embrace this new knowledge, the concepts of core (genes present in all genotypes), dispensable (genes which do not exist in all genotypes) and pan-genomes (all genes in a given species) were developed (Tettelin et al. 2005). The existence of dispensable genomes has also been reported in various crop species in recent years. Recent reports show that dispensable genes can account for a large proportion of the genes found in a species. For example, they account for about 43% in rice (*Oryza sative* L.) (Zhao et al. 2018), 50% in maize (*Zea mays* L.) (Hirsch et al. 2014; Jin et al. 2016), 36% in bread wheat (*Triticum aestivum* L.) (Liu et al. 2016b; Montenegro et al. 2017) and 38% in barley (*Hordeum vulgare L.*) (Ma et al. 2019). Clearly, with the increase in the numbers of genotypes sequenced, the percentage of dispensable genes in a species will get larger. Available results indicate that, compared with those in the core genomes, higher proportions of dispensable genes encode tolerance to biotic and abiotic stresses in various species. This seems to be the case in many species including rice (Zhao et al. 2018), wheat (Appels et al. 2018; Montenegro et al. 2017), tomato (Gao et al. 2019a), canola (Dolatabadian et al. 2019; Samans et al. 2017), soybean (McHale et al. 2012), and barley (Ma et al. 2019).

Benefiting from the decreasing cost of sequencing and the improved assembly technology, it has been feasible to construct comprehensive pan-genomes of plants with relative small and simple genomes, such as rice (Wang et al. 2018b; Zhao et al. 2018) and *Arabidopsis thaliana* (1001 Genomes Consortium 2016). However, it is still not practical to obtain deep sequences from thousand-scale genotypes for those species 3 and with large and complex genomes. Barley has a large genome (about 5.1 Gb) (International Barley Genome Sequencing Consortium 2012) with extreme abundance of repetitive elements (Wicker et al. 2017). For this

reason, the international barley community had for many years focused on obtaining a high-quality reference genome assembly based on the genotype Morex (International Barley Genome Sequencing Consortium 2012; Mascher et al. 2017), and high-quality sequences are planned for only a small number of genotypes in the on-going barley pan-genome project (Monat et al. 2019b).

Although high-quality genome assemblies are still difficult to obtain from large numbers of genotypes for species with large and complex genomes, sequences based on genotype-by-sequencing (GBS) have become available for more than thousands genotypes in some species in recent years (Juliana et al. 2019; Lu et al. 2015). In barley, GBS sequences from more than 22,000 genotypes have recently been made available (Milner et al. 2019). Together with the several available draft genome sequences (Dai et al. 2018; Liu et al. 2020; Zeng et al. 2015), the large amounts of GBS sequences offer the potential to dramatically enhance our capacity to understand the pan-genome of this important crop species.

Based on the use of GBS sequences from a large number of inbred lines, Lu et al. (2015) developed a TASSEL Pan-genome Atlas pipeline for constructing high-resolution maize pan-genome sequence anchors. GBS tags (short sequences generated by *k*-mer analysis) were firstly mapped with a high-density SNP matrix in this pipeline and the mapping results were subsequently used to train machine learning models to predict physical positions of all GBS tags. Inspired by the maize work, we identified barley pan-genome sequence anchors based on the available GBS data using genetic mapping and machine learning, and the obtained results are reported in this publication.

6.2 Methods

6.2.1 Data source

Raw GBS sequence reads from 1,140 wild barley (Table S6.1) (*H. valgure ssp. Spontaneum (k. Koch) Thell*), and 11,166 domesticated barley (Table S6.2) were extracted from the European Nucleotide Archive (ENA). Compromising on computing limitation of our machine, the domesticated panel was randomly selected and covered half-size of the IPK dataset. All raw GBS sequences were trimmed using SolexaQA ++ v3.1.3 (Cox et al. 2010) with Phred quality

threshold of 30 and read length threshold of 64 bp. The imputed SNP matrices and phenotypic data of the wild and domesticated genotypes (Milner et al. 2019) were downloaded from https://doi.org/10.5447/IPK/2018/9.

6.2.2 Source code

Perl scripts developed in this study was uploaded to LDM-RF repository: https://github.com/ShawnGao911101/LDM-RF. Data cleaning, random forest model training and result visualization were processed with Python 3.6 and the code was archived in the LDM-RF repository as well.

6.2.3 Generating unique record (UR)

KMC (v3.0.0) (Deorowicz et al. 2015) was used to generate raw GBS tags from the trimmed reads for every genotype with a *k*-mer size of 64 and occurrence no less than 30 (Fig. S6.1a). Presences of a GBS tag in all genotypes were documented into a 'Unique Record'. Each UR represents all GBS tags which are present in the same subset of genotypes (Fig. S6.1b). All *k*-mer-analyzed results were performed by script: 'create_unique_records.pl' to create an integrated UR hash and split UR hashes (W dataset: 1500 URs per hash; D dataset: 2000 URs per hash) for parallel computing. The one-to-many links between UR and GBS tags were retrieved from the integrated hash with 'creat_index.pl'.

6.2.4 Linkage disequilibrium mapping (LDM) of UR against SNP matrix

Tassel v5.0 was employed to convert the SNP matrices from VCF to the Hapmap format (Bradbury et al. 2007). Script 'calculate_sum.pl' was used to calculate and descending sort the 'SUM' values for every UR in split hash against the whole SNP matrix (Fig. S6.2). 'SUM' is designed as a Cohen's Kappa statistics (Cohen 1960) for evaluating the agreement between SNP genotypes and presences/absences of a UR in all of the barley genotypes. A result file was generated and saved for every UR, including associated SNPs and their physical positions. Associated SNPs were defined as: starting with the highest SUM value and stopping once the SNPs on all seven chromosomes and the assembly of unknown reads appears in the descending sorted results. The position of an associated SNP with the highest SUM value was taken as the LDM position of the UR.

6.2.5 Selection and training of machine learning model

Three machine learning methods, the artificial neural networks (ANNs), the support vector classifier (SVC), and the random forest (RF), are investigated on the position confidence evaluation of GBS tags (See details in the Fig. S6.3 and S6.4). Through the comparison, RF model presented the best performance and, therefore, was finally selected to classify the GBS tags in this study. The scikit-learn module (Pedregosa et al. 2011) in Python along with the Jupyter notebook was used for training RF models. BWA (Version 0.7.15) mem (Li 2013) was used to align all GBS tags to the Morex genome and the tags uniquely and perfectly alignment were defined as UAMTs. The coordinates of the best hit were taken at the physical positions of the corresponding GBS tags.

Referring to the model training result in maize (Lu et al. 2015), UAMTs were defined as four classes based on the distance (Dist) between the physical and LDM positions: Class 1: Dist ≤ 1,000 bp; Class 2: 1000 bp < Dist ≤ 100 Kbp; Class 3: 100 Kbp < Dist ≤ 10 Mbp; and Class 4: Dist ≥ 10Mbp. Such a grouping method allows the trained RF classifiers predicted GBS tags on different levels of accuracy. All positions were transformed according to the equation: transformed position = chromosome × 1E9 + position. Classification accuracies were evaluated using the random forest out-of-bag cross-validation on the training dataset. Taking the tag class as dependent variable, we randomly selected 200,000 UAMTs as training dataset (25% of them as test set) for adjusting model parameters and features used in establishing RF classifiers. Finally, a total of ten features (Table S6.3) were calculated for each UR based on the LDM result using script: 'creat_ML_features.pl'. The final training parameters were n_estimators = 90; min_samples_leaf = 2; max_features = 6 and the other parameters were kept as default. Receiver operating characteristic (ROC) curves and precision-recall Curves were plotted for each of the tag Classes to evaluate the robustness of RF models. The final RF classifiers were built using the entire UAMT datasets and they were taken to predict the Classes of class-unknown tags. In the study reported here, we focused on results of Class 1 to 3, which were defined as reliable tags.

6.2.6 Identified presence/absence variation (PAV) tags

Within the tags identified as Classes 1 to 3, those absent from the reference genome of Morex were defined as PAV-I tags. The tags whose LDM positions were 100 Mb away from every hit of their physical positions were defined as PAV-II tags.

6.2.7 Validation of PAV-I tags

To validate the credibility of PAV-I tags, we used BWA mem to align them against the genome of the hulless barley genotype (Zangqing320) (Dai et al. 2018) and filtered out those with perfect alignment (ZQ-tags). We then aligned 2-Kb flanking sequences of Zangqing320 around ZQ-tags back to the Morex genome. If the flanking sequence aligned within 10 Mbp of the LDM position of a given ZQ-tag, it was treated as a successful validation (Fig. S6.5).

6.2.8 Genome-wide association studies (GWAS) between the GBS tags and phenotypic traits

GWAS was performed using GAPIT v2.0 (Tang et al. 2016) with the compressed mixed linear model. Phenotypic data included flowering time, plant height and thousand grain weight. The domesticated genotypes that have phenotypic records (6,073 of the 11,166 accessions) were analysed in GWAS against the domesticated SNP matrix which contains 76,102 SNPs (Table S6.4). As the relative position of PAV tags were represented by the corresponding SNPs, we directly employ the entire SNP matrix in GWAS. Quantile-Quantile (Q-Q) plots were generated between the significance of SNPs associated with PAV tags (PAV SNPs) and Non-PAV tags (Non-PAV SNPs) to evaluate the relative phenotypic influence from these two types of GBS tags. SNPs with minor allele frequency (MAF > 0.095) were selected to confirm whether PAV and Non-PAV SNPs had similar MAF distribution with 0.25 as median (Fig. S6.6).

6.3 Results

6.3.1 General introduction of the pipeline used in this study

To build a framework for the barley pan-genome, linkage disequilibrium mapping (LDM) were used in conjunction with machine learning (ML) algorithms to anchor GBS tags from 11,166 domesticated (D dataset) and 1,140 wild (W dataset) barley genotypes to the Morex reference genome coordinates (Fig. 6.1). Through K-mer analysis (Fig. S6.1), a total of 18.0 million and 9.8 million raw GBS tags were obtained from the D dataset and W dataset, respectively. According to whether a tag had a unique and perfect hit on the Morex genome, these raw GBS tags were separated to uniquely aligned to the Morex tags (UAMTs) and tags waiting for classification. All raw GBS tags were linkage disequilibrium mapped to the SNP matrix and the position of the most significant SNP was taken to be the LDM position of the tag (Fig. S6.2). In order to obtain high-quality pan-genomic anchors, the UAMT dataset were used to train ML models to classify the other raw GBS tags and those tags passed the model selection were identified as sequence anchors. To select the most suitable algorithms, three popular methods, including random forest (RF), supporting vector classifier (SVC) and artificial neural networks (ANNs), were tested on UAMTs from both datasets with ten features (Table S6.3). The comparison results (Fig. S6.3 and S6.4) shows that the RF model performed best in terms of classification accuracy of each tag class. Therefore, RF was employed as the final algorithm to training model and classifying GBS tags for both D and W datasets.

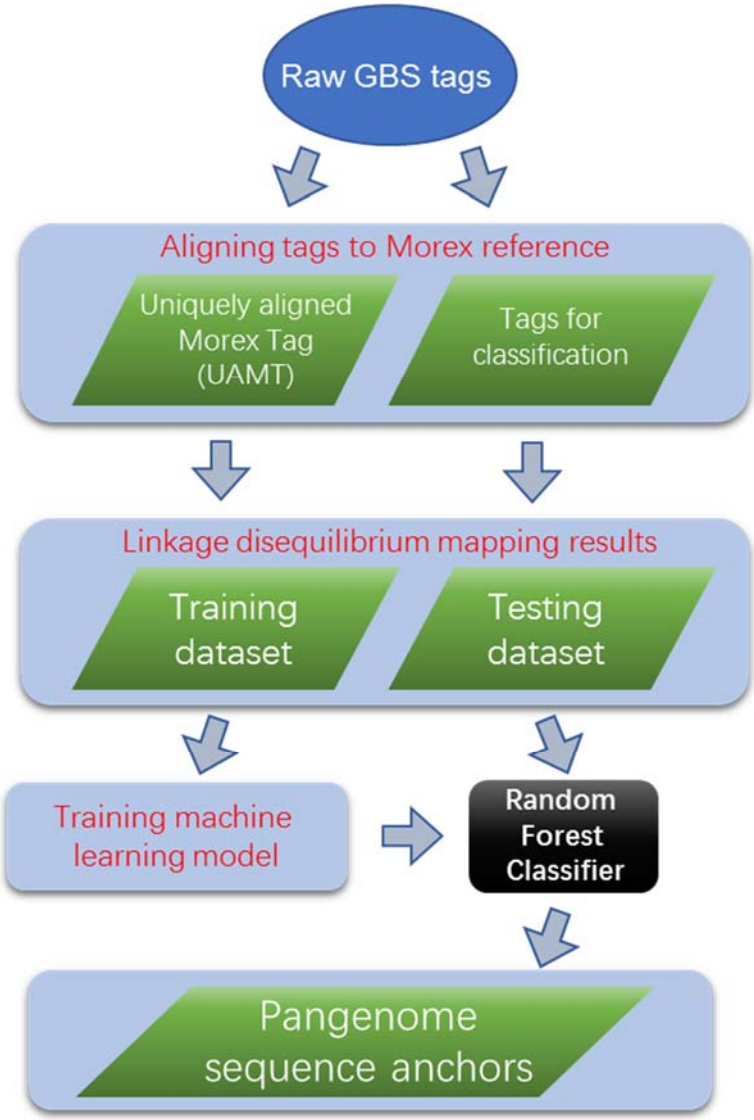


Fig. 6.1 Generating barley pangenome sequence anchors using a strategy combining linkage disequilibrium mapping and machine learning.

6.3.2 Classification accuracies of RF classifiers

Of the raw GBS tags obtained, 1.5 million from the D dataset and 1.2 million from the W dataset were UAMTs. We randomly selected 200,000 of the UAMTs to train RF classifiers and test their classification accuracies. The ten features showed various degrees of importance in the training processes for both models (Fig. S6.7). Though the distribution of UAMTs strongly skewed to Class 4 (Table S6.5), both classifiers showed high level of accuracies for every class (Fig. 6.2). The classifier for those GBS tags from the D dataset performed better than those from the W dataset on classification accuracies for Classes 2 and 3. This was largely due the

different size between the domesticated panel and the wild panel. The much larger number of domesticated genotypes used should provide more precise linkage disequilibrium mapping (LDM) results to train more robust ML model.

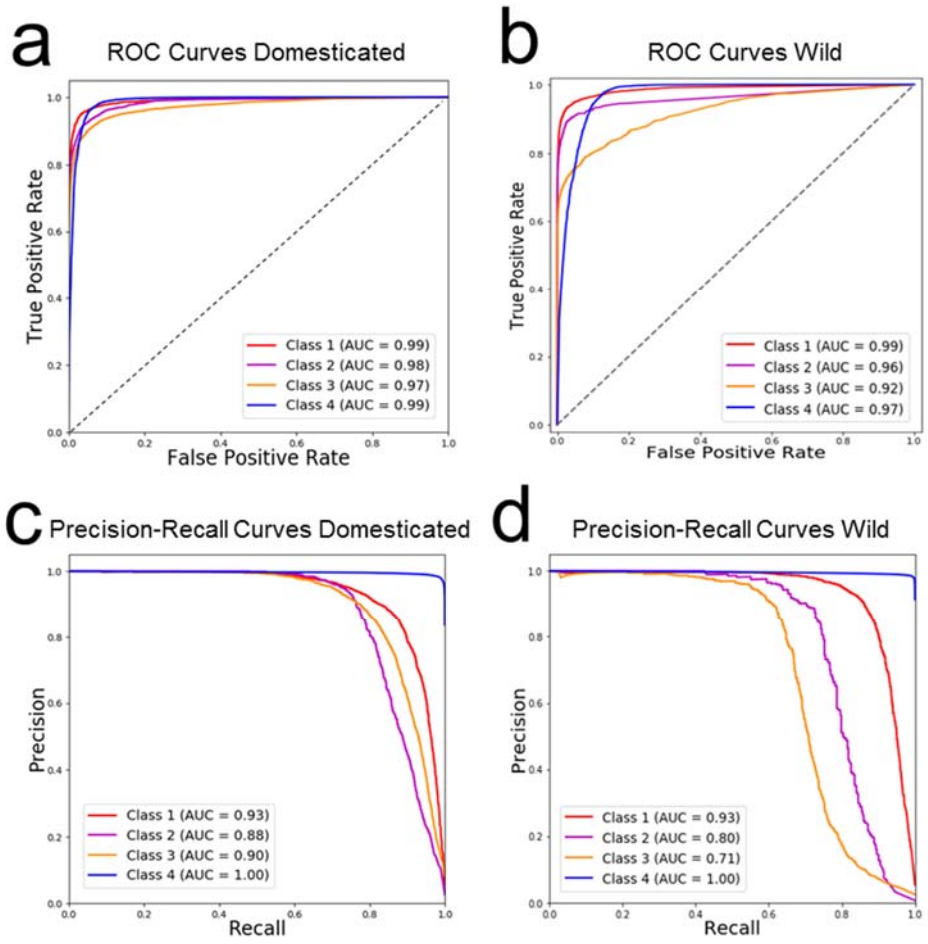


Fig. 6.2 Receiver operating characteristic (ROC) and precision-recall curves showing classification accuracies for four tag classes of random forest classifiers. (a) and (b) are ROC curves for the D and W dataset, respectively. (c) and (d) are precision-recall curves for the D and W dataset, respectively.

6.3.3 Classification results of RF classifiers

The final RF models, trained with all UAMTs, were applied to all LDM results of both the domesticated and wild GBS tags. A total of 1,380,465 domesticated and 443,871 wild GBS tags were categorized into three different classes (1, 2 and 3), which were considered as pan-genome sequence anchors from hereon. Among them, about 66.8% of those from the D dataset and 83.4% of those from the W dataset were Non-PAV tags (Fig. 6.3a and c) as their physical

positions on the Morex genome were the same as the LDM positions (Fig. 6.3b and d). These results provide further evidence suggesting the robustness and credibility of the RF classifiers.

The two classifiers successfully sorted part of the Morex-absent tags into reliable tags and they were defined as PAV-I tags. These tags (Fig. 6.3a and c) accounts for 17.9% of those from the D dataset and 6.4% of those from the W dataset. There was another type of tags whose LDM positions were at least 100 Mbp away from the corresponding physical positions. We named them as 'PAV-II' tags which were present on the Morex genome but absent in some of the other genotypes used here. PAV-II tags (Fig. 6.3a and c) accounted for 15.4% of those from the D dataset and 10.2% of those from the W dataset.

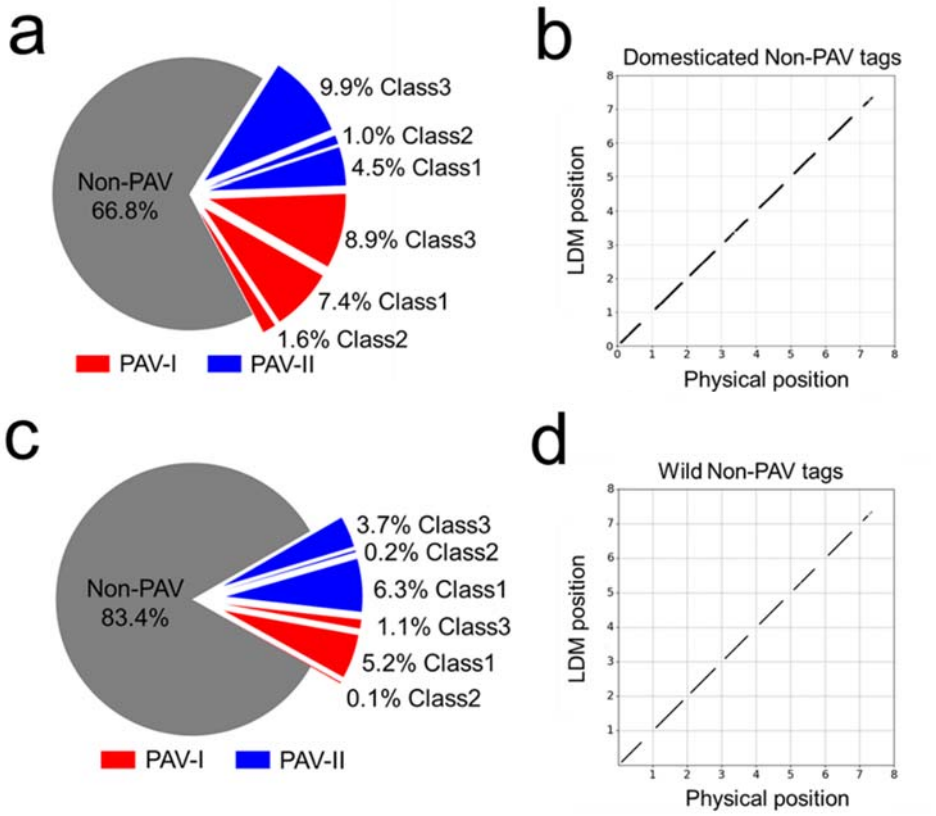


Fig. 6.3 Classification results of random forest classifiers. The pie charts show the proportions of three types of tags from (a) domesticated dataset and (c) wild dataset. The scatter plots show the physical positions against linkage disequilibrium mapping (LDM) positions of the Non-PAV tags from (b) the domesticated and (d) wild datasets. The numbers on x and y axis indicate barley chromosomes ('8' indicates the unknown chromosome). All positions are transformed using the equation: transformed position = chromosome × 1E9 + position.

6.3.4 Genome-wide distribution of the PAV tags

The PAV-I tags from the D and the W datasets shared a similar chromosome-wide distribution pattern (Fig. 6.4a and c). With a few exceptions, most of these tags located on the distal regions of a chromosome and only a few were found in the pericentromeric region. The distributions of PAV-II tags were similar to those of PAV-I, but the densities of the former in the distal regions were even higher for most of the chromosomes (Fig. 6.4b and d) for data from both the D and the W datasets. Interestingly, few of the PAV-II tags from the D dataset were located in the pericentromeric region on chromosome 1H (Fig. 6.4b), which was different for those tags from the W dataset (Fig. 6.4d).

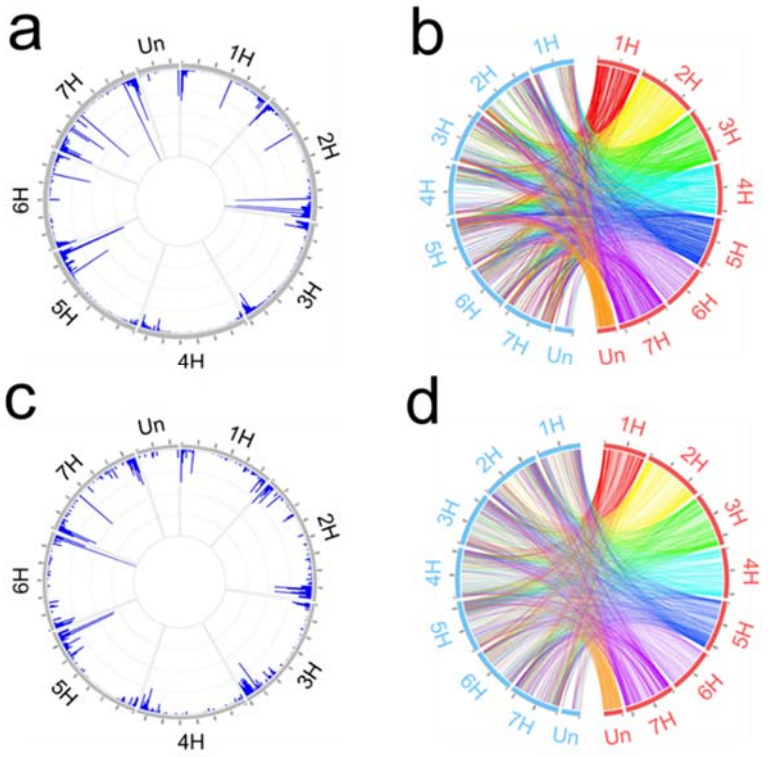


Fig. 6.4 Genome-wide distribution of PAV tags. (a) Domesticated PAV-I tags. The highest bar represents 10,000 tags; (b) Domesticated PAV-II tags. The physical and LDM positions are shown in red and cyan, respectively; (c) Wild PAV-I tags. The highest bar represents 1,200 tags; (d) Wild PAV-II tags. The physical and LDM positions were showed in red and cyan, respectively.

6.3.5 Validation of the PAV-I tag positions

To validate the locations of the PAV-I tags, they were examined against the genome sequences of the barley genotype Zangqing 320. Of the tags absent in the Morex genome, 1,823 from the domesticated and 623 from the W dataset were perfectly and uniquely mapped on the Zangqing320 genome. For these tags, 83.6% of domesticated tags (Table S6.6) and 88.6% of wild tags (Table S6.7) had their best hit within 10 Mb around the LDM positions of the corresponding PAV-I tags. These results confirmed the credibility of the locations of PAV-I tags.

6.3.6 Comparative effects of PAVs and non-PAVs on phenotypic variations

To evaluate if there were differences in phenotypic variation between PAVs and Non-PAVs, genome-wide association studies were conducted for three traits (flowering time, plant height and thousand kernel weight) on 6,073 domesticated barley genotypes. As PAVs were genetically mapped by co-segregation with SNPs, the PAV-associated SNPs can be used as agents to assess their genetic effects. The SNPs where PAV tags mapped were defined as PAV SNPs (including PAV-I SNPs and PAV-II SNPs). The SNPs that did not associate with PAV tags were termed as Non-PAV SNPs. Results from this analysis showed that these two types of SNPs contributed differently to the three different traits. Compared with those Non-PAV SNPs, PAV SNPs showed much stronger effect on FT (Fig. 6.5 a and b). However, those Non-PAV SNPs explained more variations for both PH (Fig. 6.5 c and d) and TGW (Fig. 6.5 e and f).

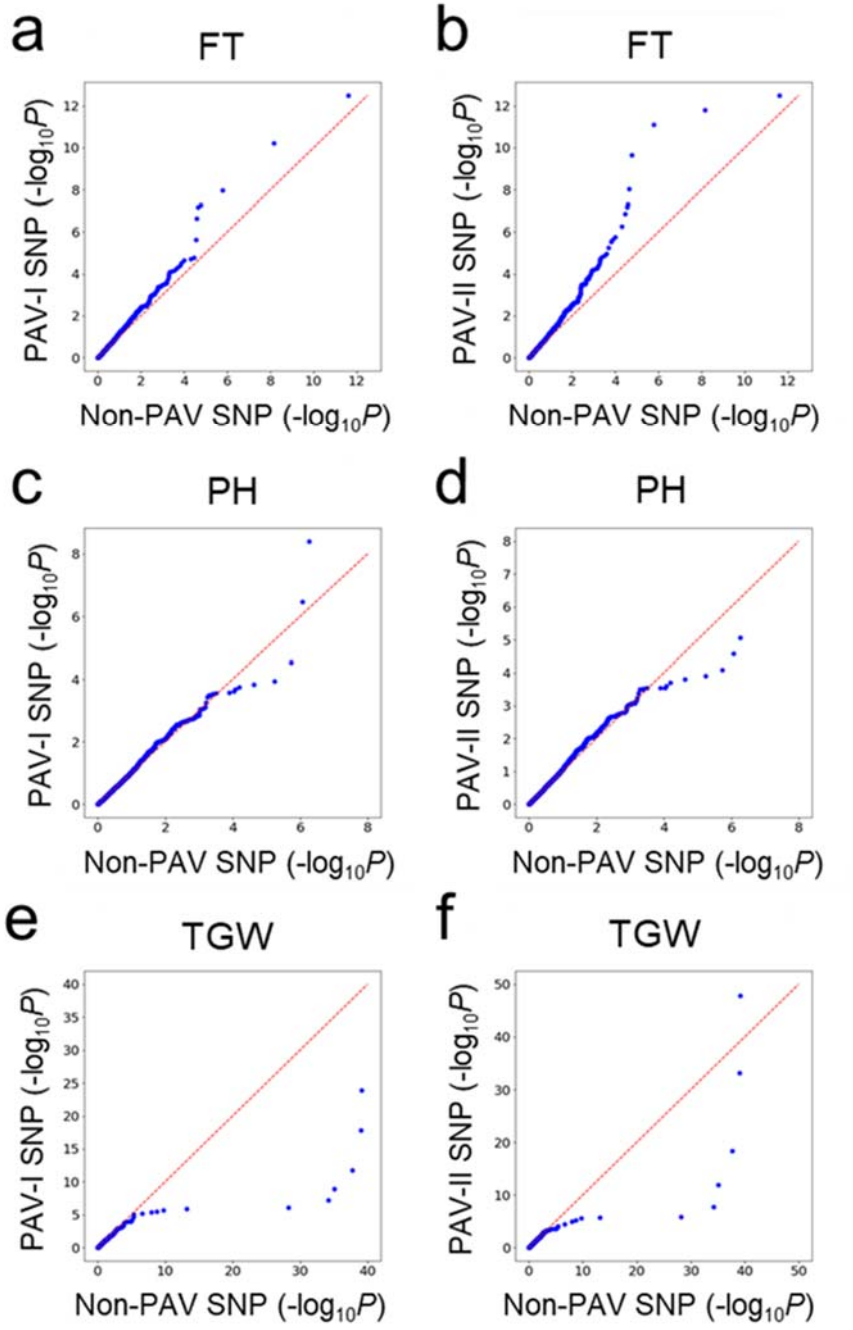


Fig. 6. 5 Q-Q plot of pairwise comparison of GWAS hits between Non-PAV SNPs and PAV SNPs. (a, c, e): Non-PAV SNPs vs PAV-I SNPs. (b, d, f): Non-PAV SNPs vs PAV-II SNPs. 'FT' stands for flowering time; 'PH' stands for plant height; 'TGW' stands for thousand grain weight.

6.4 Discussion

Using genetic mapping and machine learning, we generated an informative dataset of barley pan-genome sequence anchors consists of over 1.38 million of GBS tags from 11,166 domesticated genotypes and 443,871 GBS tags from 1140 wild barley genotypes. Among them, 532,253 were PAV tags. Based on syntenic relationships with the sequences of the hulless barley genotype Zangqing320, 83.6% of the PAV tags from the domesticated genotypes and 88.6% of those from the wild barley genotypes were correct. Association analyses against flowering time, plant height and thousand grain weight showed that PAVs contributed significantly to each of these characteristics. However, the relative contribution of PAV and non-PAV tags varied among these traits.

Previous studies have shown that the barley genome is characterized by very low genetic diversity and recombination rate in the pericentromeric regions on each of its seven chromosomes (Liu et al. 2020; Mascher et al. 2017). Thus, it was anticipated that the distributions of the PAV tags along each of the barley chromosomes could be uneven. However, compared with those for either the genetic diversity or recombination rate, the distributions of the Morex-absent GBS tags are more extreme. More than 90% of the tags located in less than a quarter of the regions at the distal ends on each of the chromosomes. One of the possible contributing factors for the extreme distributions of the GBS tags is that machine learning used in this study was based on association analysis which heavily relies on genetic recombination. The low genetic diversity compounded by the low recombination rates (Künzel et al. 2000; Mascher et al. 2017) mean that it would be even harder for those tags located in the pericentromeric regions to get over to the same cut off used to determine reliable GBS tags in this study. Another potential factor is that the restriction enzymes, PstI-MspI (Milner et al. 2019), used for generating GBS data may result in fewer sequenced fragments in the pericentromeric region of each chromosome (Ott et al. 2017).

The total number of unique GBS tags obtained from the cultivated genotypes is much larger than those from the wild genotypes. However, the real difference in diversity between the wild and domesticated genotypes is obscured by the difference in the numbers of genotypes used. When calculated on a per genotype basis, the average number of the reliable GSB tags per genotype detected from the wild genotypes is more than three times larger than those detected

from the cultivated genotypes. The huge difference in relative diversity detected in this study between wild and cultivated genotypes was not unexpected. Previous studies showed that wild barley genotypes are a rich source of genetic variation for breeding programmes as less than half of the alleles found in wild barley could be detected in cultivars (Ellis et al. 2000; Tombuloglu et al. 2015). The results obtained in this study provide further evidence showing that wild barley genotypes do not only provide a rich source of new alleles but also novel genes for breeding programs as large numbers of genes which may have been lost during evolution, domestication or breeding (Liu et al. 2020; Ma et al. 2019).

It is of interest to note that the relative contributions between the PAV and non-PAV tags varied among the three assessed characteristics assessed in this study. The GWAS results showed that PAV tags made the most significant contribution to the phenotypic variance of flowering time. One of the possible reasons for the difference is that the three characteristics have different evolutionary significance. Recent reports show that structural variation plays important roles in ecogeographical adaption of many crops (Gabur et al. 2019; Schiessl et al. 2019), such as rice (Zhao et al. 2018), wheat (Appels et al. 2018; Montenegro et al. 2017), tomato (Gao et al. 2019a), canola (Dolatabadian et al. 2019; Samans et al. 2017), soybean (McHale et al. 2012), and barley (Ma et al. 2019). Among the three characteristics assessed in this study, flowering-time is likely the most critical factor affecting crop adaption. The importance of flowering time in plant adaption has been repeatedly reports in earlier studies (Diaz et al. 2012; Würschum et al. 2015).

6.5 Conclusion

In this study, we identified 1.8 million barley pan-genome sequence anchors covering 11,166 domesticated and 1,140 wild barleys. A total of 532,253 presence/absence variations tags were identified and they provided an overview of genome-wide PAV distributions in barley. Our findings present that PAV could significantly influence barley flowering-time. These PAV tags would assist various genetic studies including identification of structural variation, genetic mapping and marker developments in barley. Moreover, these pan-genome sequence anchors could help to build the comprehensive barley pan-genome through positioning the scaffolds/transcripts which could not be placed by computing alone.

6.6 Abbreviations

ANNs: Artificial Neural Networks

D: Domesticated

GBS: Genotype-by-Sequencing

GWAS: Genome-Wide Association Study

LDM: Linkage Disequilibrium Mapping

MAF: Minor Allele Frequency

ML: Machine Learning

PAV: Presence/Absence variation

RF: Random Forest

ROC: Receiver Operating Characteristic

SNP: Single Nucleotide Polymorphism

SVC: Supporting Vector Classifier

UAMT: Tags Uniquely Aligned on the Morex genome

UR: Unique Record

W: Wild

6.7 Availability of data

Non-PAV, PAV-I and PAV-II GBS tags from both the domesticated and wild genotypes are available at https://github.com/ShawnGao911101/LDM-RF.

File name: GBS_tags_collection_Gao_etal_2019.7z

Chapter 7 General conclusion

Fusarium crown rot (FCR), mainly caused by *Fusarium pseduograminearum*, is a chronic disease for barley and wheat in semi-arid areas worldwide. Although some farming practices, including crop rotation, stubble burning and tillage management, can decrease the incidence of FCR to some extant (Cook 1980, 2001), these practices may not always economical, practical and friendly for the environments. It has been recognized that breeding and growing resistant varieties are a key component in effectively controlling FCR. Facilitating the efforts of breeding FCR resistant varieties forms the main objective of my PhD study.

Characterizing a new source of FCR resistance forms the first part of my study, leading to the mapping of novel QTL conferring FCR resistance on the chromosome arm 6HL. This locus could explain up to 28.3% phenotypic variance of disease severity in the mapping population and reduce FCR magnitude by 22.9% in another validation population (Chapter 3). These results once again demonstrated that, although it is widely believed that genetic control to diseases caused by necrotrophs tend to be highly quantitative, loci with large effects do exist. The existence of such loci would make it feasible to breeding varieties with high levels of resistance.

It has long been recognized that QTL mapping can only provide limited resolution (Paterson et al. 1988). One of the main reasons is that the segregation of undesirable characteristics would affect the accuracy of phenotypic evaluation on targeted trait. Thus, loci detected from QTL mapping can only be treated as putative. Their true values need to be validated. Near isogenic lines are widely used for such validation. For these reasons, the second part of my PhD research was to validate one of the earlier reported loci. This was the one located on the chromosome arm 1HL (Chen et al. 2013b). To validate this locus, we developed and assessed five pairs of near-isogenic lines (NILs) and obtained transcriptomic data from three of these NIL pairs (Chapter 4). We did not only validate the effects of this locus by analyzing the NILs, but also delineated it into an interval of 11.0-Mbp. A set of candidate genes within the region were identified and functional analysis found that the resistance controlled by this locus can be highly novel and are likely different from the possible mechanism reported previously (Habib et al. 2018). Clearly, genes underlying the locus have to be cloned before the exact mechanism of its resistance can be clarified.

Molecular markers are powerful tools in enhancing breeding efficiencies. The values of markers are, to a large degree, determined by their close linkage with their targeted loci. To develop user-friendly markers which can be used to reliably trace a certain locus in breeding programs, we developed a population consisting of 1,180 lines derived from one of the NIL pairs targeting the 1HL locus. Six markers co-segregating with this locus were developed within a genomic region spanning 487 Kb. In addition, three candidate genes carrying non-synonymous single nucleotide polymorphisms (SNPs) between the resistant and susceptible alleles were also pinned down in this region (Chapter 5). These results form a solid foundation for cloning the causal gene underlying the 1HL locus. The co-segregating markers would be invaluable to effectively incorporate the locus in breeding programs.

Until a few years ago, developing a high-quality reference genome from a single genotype had been an exerted effort of the international communities for both barley and wheat. However, it has been known in the last few years that only one or even several high-quality reference genomes may not enough for identifying all species-wide genetic variations (Hirsch et al. 2014; Saxena et al. 2014; Springer et al. 2009; Zhao et al. 2018). Each genotype has its own unique genes but absent in other genotypes. Thus, constructing a pan-genome containing genes from a large number of individuals is highly desirable for both genetic study and breeding programs in a given species.

Constructing a pan-genome consisted of many chromosome-scale genomes is currently feasible for plants with small and simple genomes such as rice (Zhao et al. 2018) and *brachypodium* (Gordon et al. 2017). However, due to the financial burden and laborious computing cost, it is still not feasible to get high-quality genome assemblies for large numbers of genotypes for species with huge and complex genomes sch as barley (Mascher et al. 2017). As a pilot work of building comprehensive pan-genome of barley, we established a high-density physical map based on pan-genome anchors from GBS data of more than 12,000 genotypes (Chapter 6). This pan-genomic physical map consists of 1.8 million anchors and 532,253 of them represent present or absent variations. This work shows the outline of the barley pan-genome at first time, especially the distribution of structural variations. The physical map would not only be useful for developing markers for barley genetic studies and breeding, but also for constructing a comprehensive pan-genome in barley in the future.

Appendixes

Supporting figures

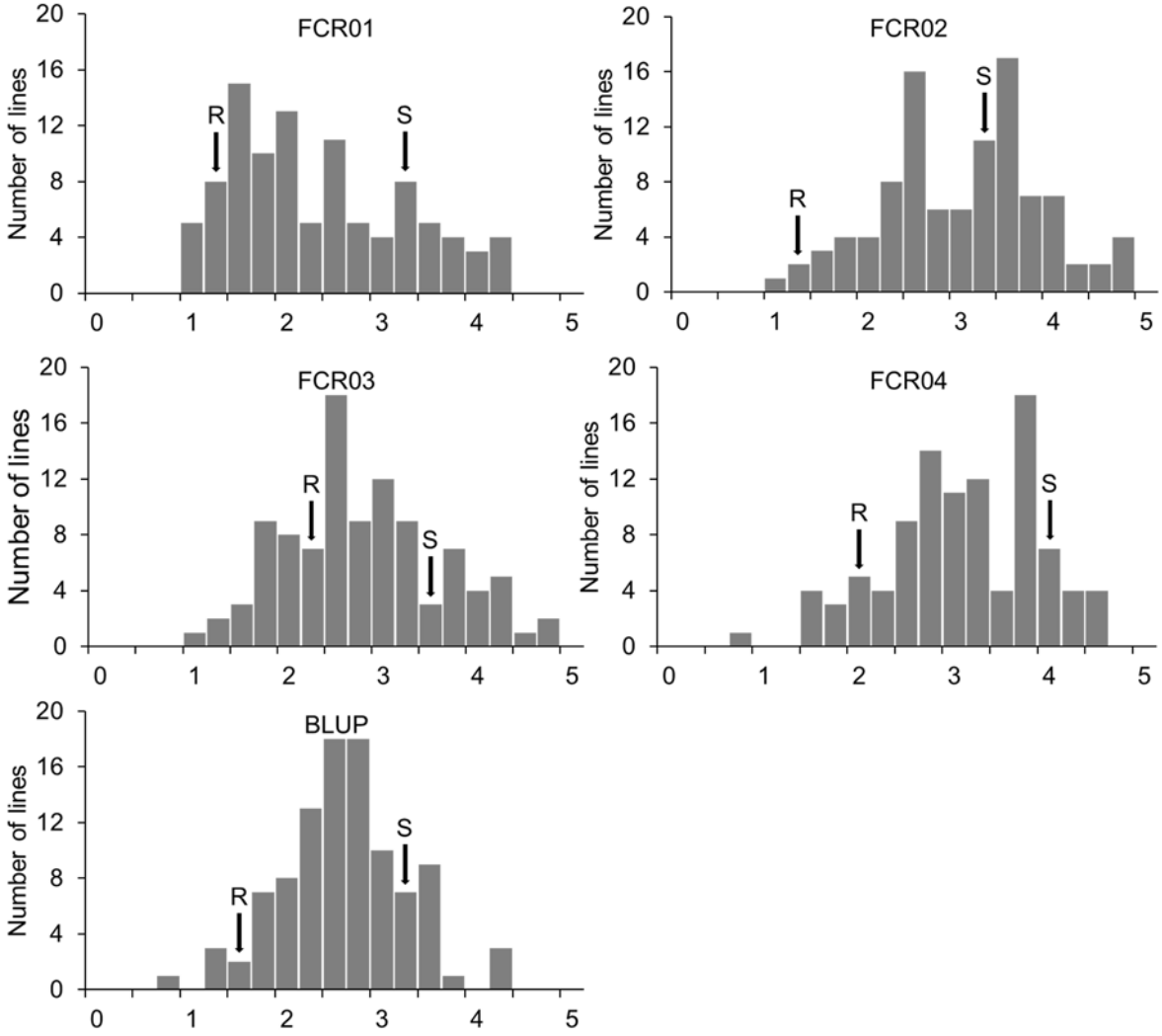


Fig. S3.1 Distribution of disease index (DI) values for the four FCR severity tests and the BLUP data from the mapping population Fleet/AWCS799. “S” stands for susceptible parent Fleet and “R” for resistant parent AWCS799.

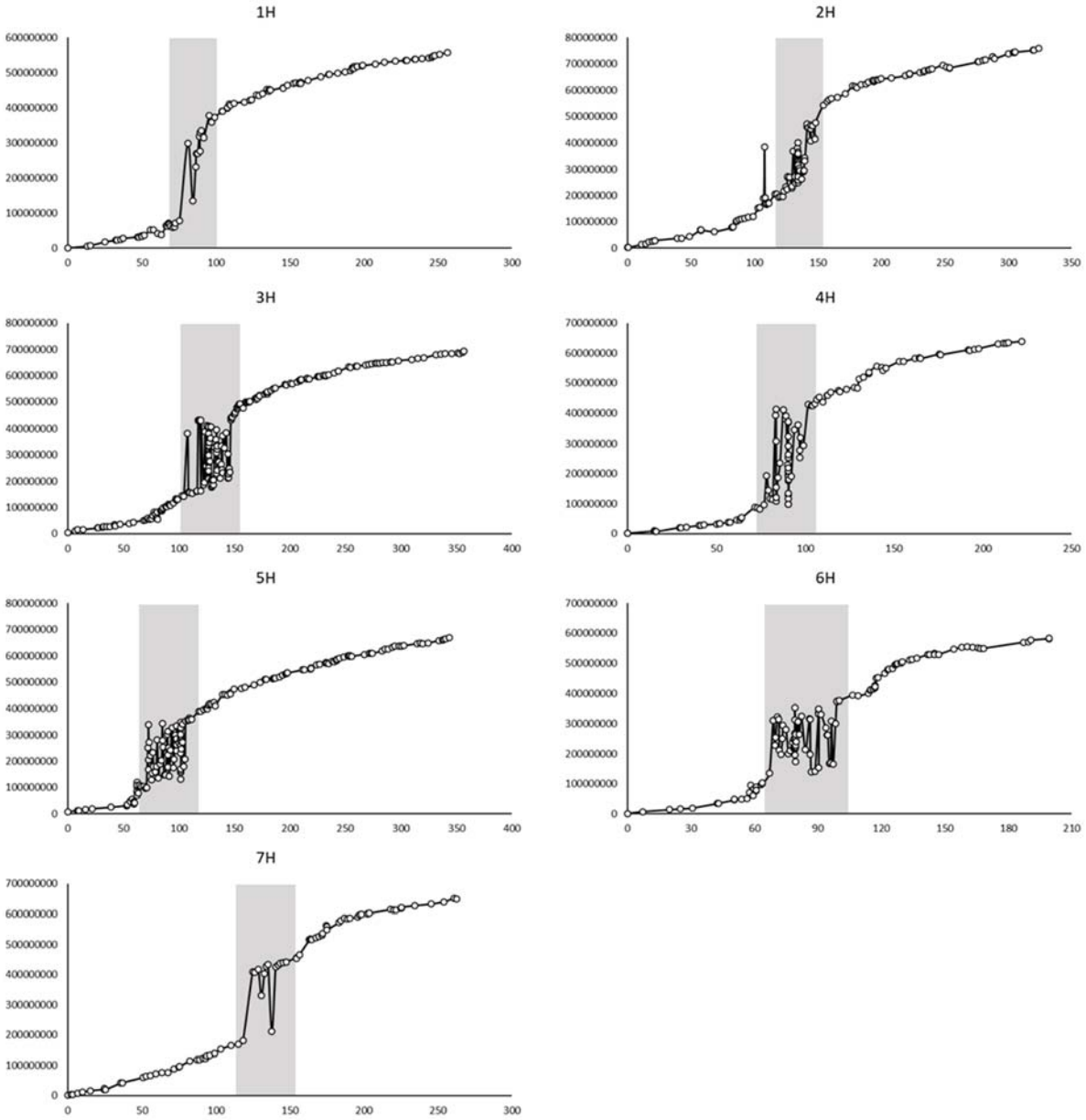


Fig. S3.2 Syntenic relationships for the mapped markers between the genetic and physical positions. Horizontal-axis represents genetic distance (cM) and vertical-axis represent physical distances (base pairs). Peri-centromeric regions were highlighted in gray shadow.

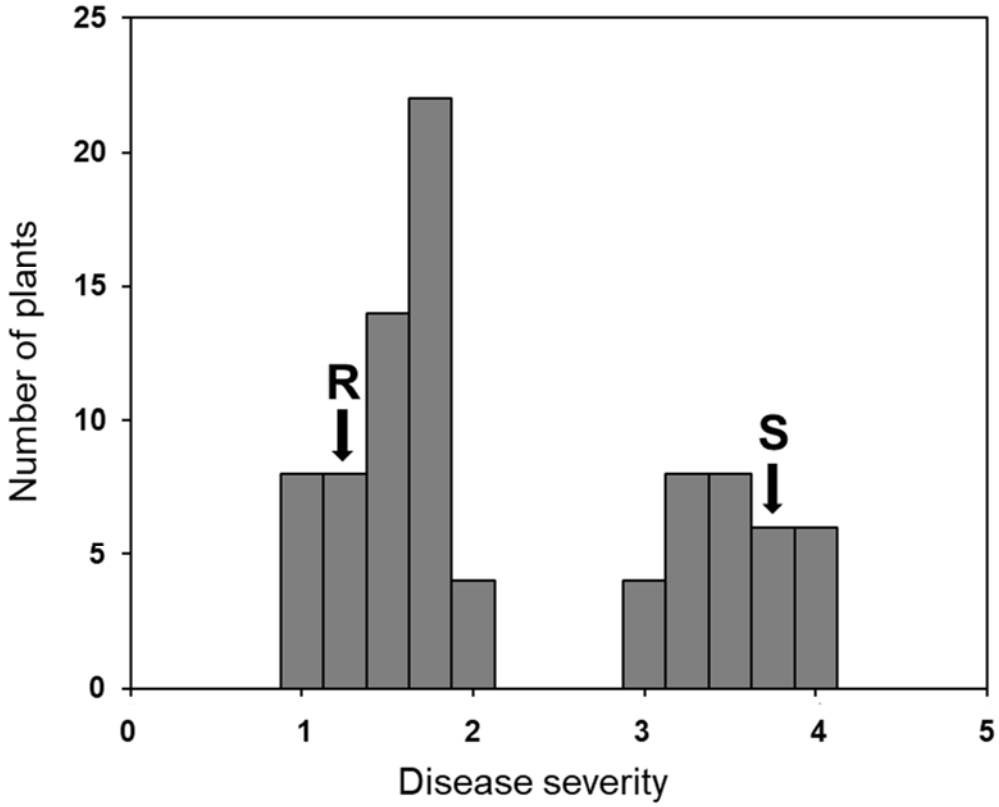


Fig. S5.1 Distribution of FCR severity in the 88-line subpopulation. The ‘R’ and ‘S’ arrows indicate the average FCR severities of the resistant and susceptible parents of the NIL-derived population.

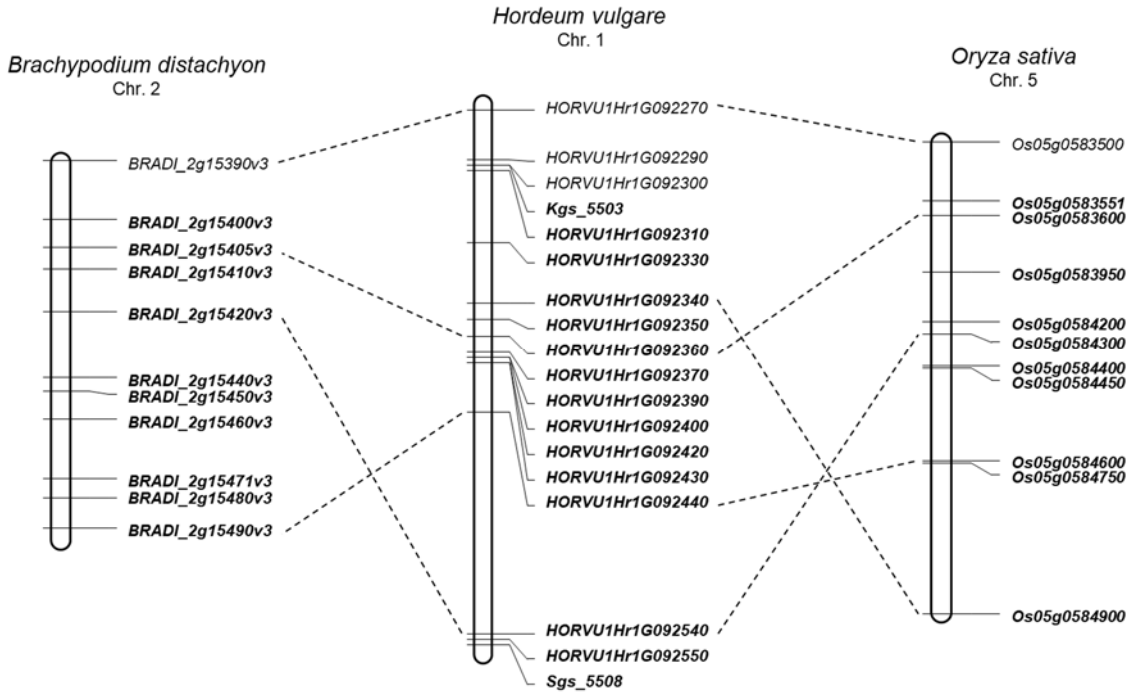


Fig. S5.2 Synteny of genomic region surrounding *Qcrs.cpi-1H* with *Brachypodium* and rice. Only high confidence genes are shown. *Kgs_5503* and *Sgs_5508* are the markers flanking the *Qcrs.cpi-1H* locus. Genes located in homoeologous regions are in bold.

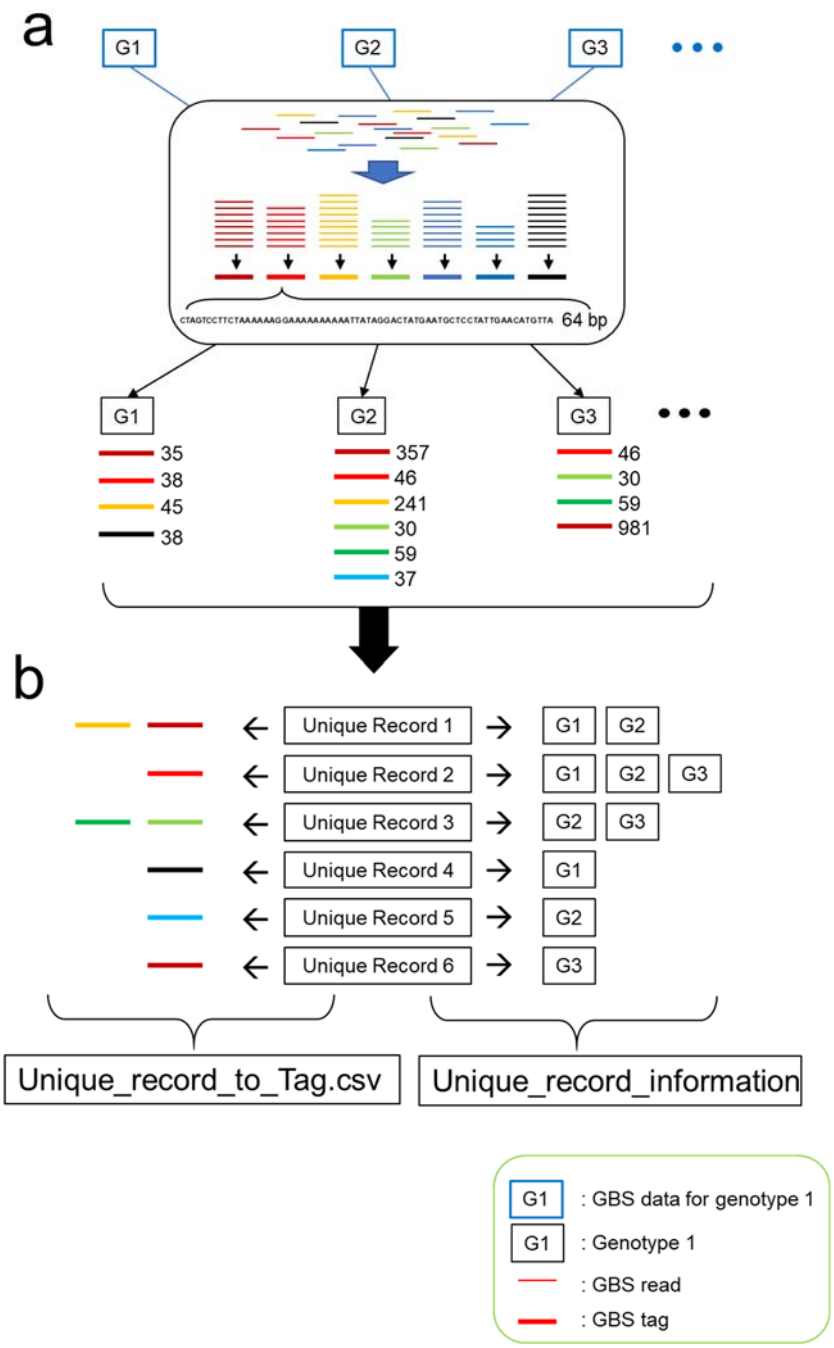
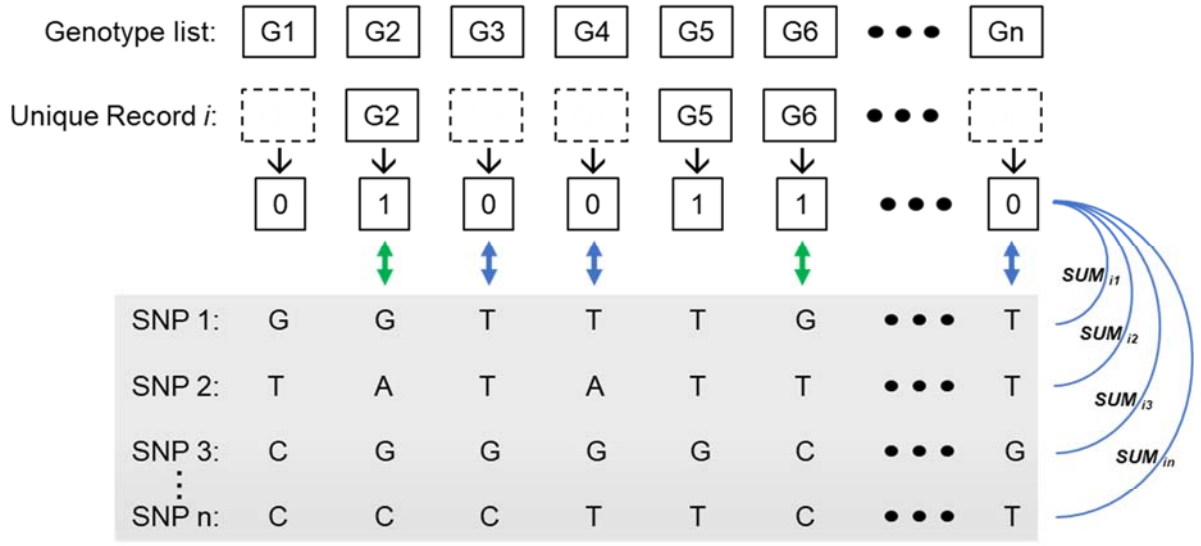


Fig. S6. 1 The pipeline used in generating the GBS tags and unique record (URs) used in linkage disequilibrium mapping. (a) K-mer analysis for each of the genotypes, and (b) Sorting GBS tags into URs.



$$SUM = \frac{M1 + M0}{T1 + T0}$$

Fig. S6. 2 Linkage disequilibrium mapping of unique records (URs) against the SNP matrix. The genotype list includes all genotypes used in the analysis. Unique Record *i* contains a subset of the genotypes. It is transformed into an '0-1' binary sequence based on the genotype list, where '0' stands for the presence of an UR in the genotype and '1' stand for absence. The equation shows how 'SUM' is calculated between UR and a single SNP, where 'M1' represents the number of matched pairs of '1' and a certain type of nucleotide and 'M0' means the number of matched pairs of '0' and the other type of nucleotide. As each SNP locus contains binary nucleotide, there will be two ways to match '0' and '1' to nucleotides. Taking the SNP 1 as example, the match will be either '1-G and 0-T' or '1-T and 0-G'. In the script 'calculate_bino.pl', we defined that the larger value generated by the two ways will be kept for each UR. Therefore, the closer the 'SUM' is to 2, the more consistent between the UR and a certain SNP is.

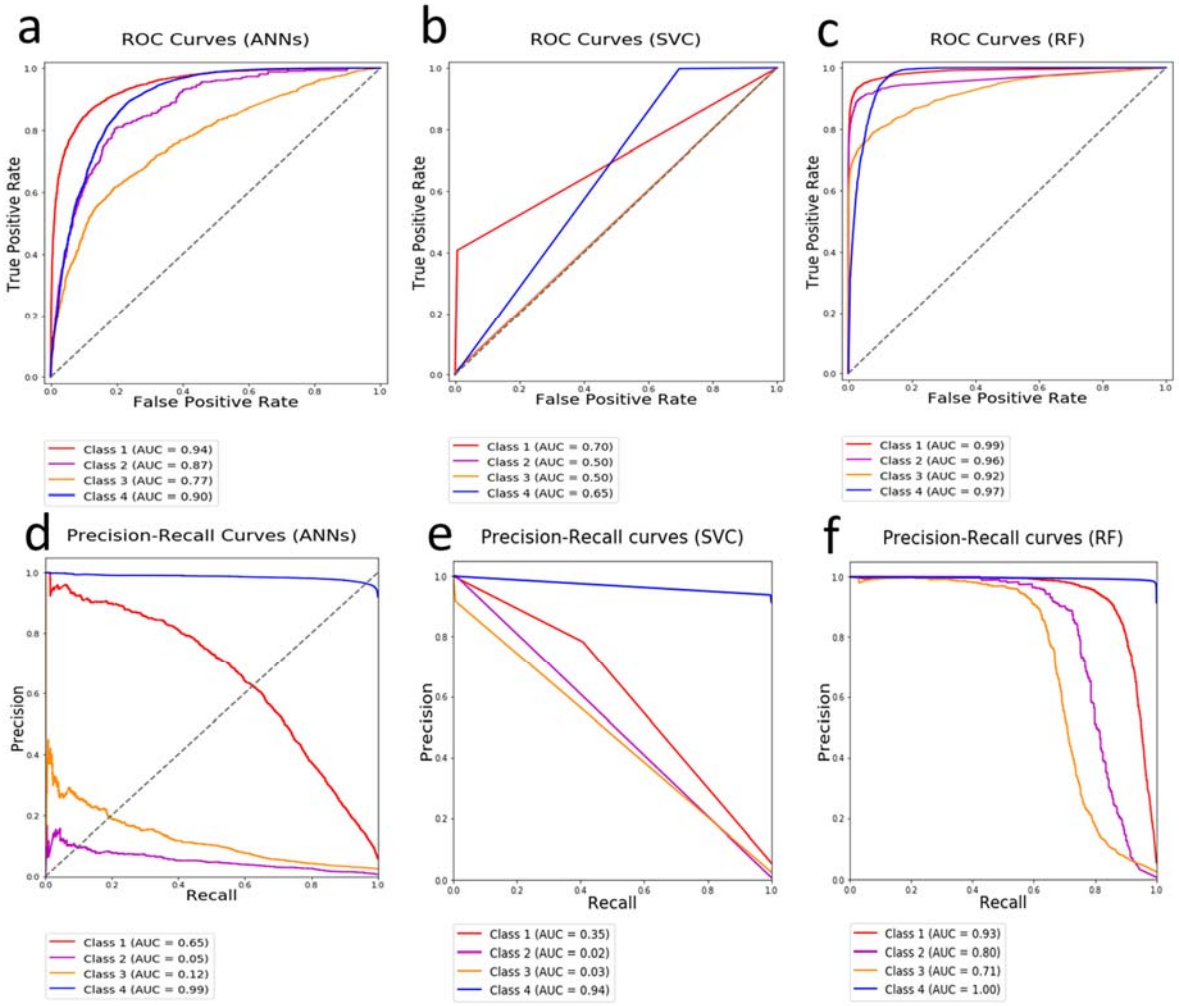


Fig. S6. 3 The machine learning algorithm comparison on the position confidence evaluation (wild dataset). As illustrated in above, the R (**c** and **f**) is superior to the ANNs (**a** and **d**) and SVC (**b** and **e**). For the ROC curve comparison, the AUC values for all four classes using the RF are significantly larger than these using the ANNs and the SVC. On the other hand, from the precision-recall curves, the difference among the RF, the ANNs, and SVC, are more remarkable, showing the RF is more efficient in the unbalanced data with high AUC values (Class 1: 0.93, Class 2: 0.80, Class 3: 0.71, and Class 4: 1.00). From above analysis, compared with the ANNs and the SVC, the RF is a more reliable machine learning method for the position confidence evaluation of the GWAS results.

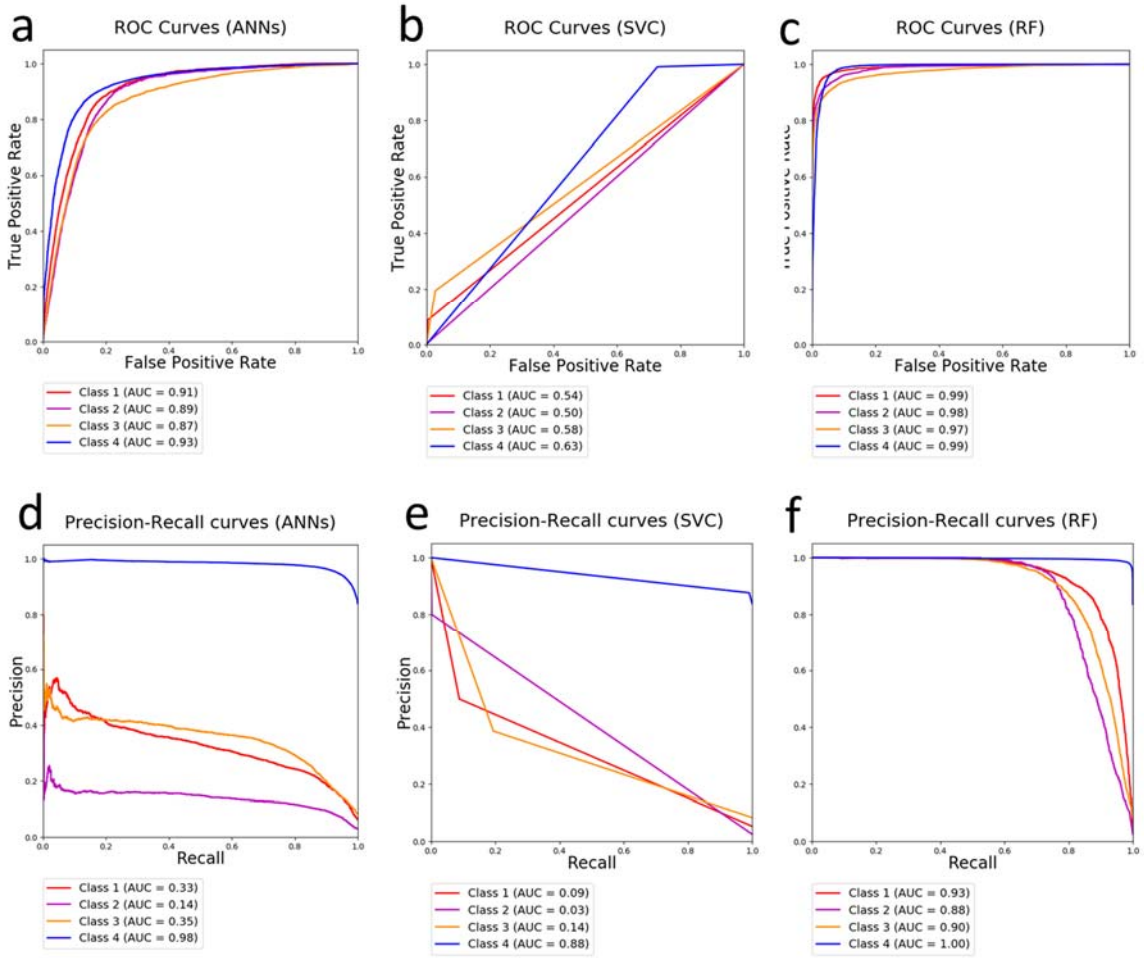


Fig. S6. 4 The machine learning algorithm comparison on the position confidence evaluation (domesticated dataset). Like the results presented in the wild dataset, the RF model (**c** and **f**) showed better performance of classification than the ANNs (**a** and **d**) and the SVC models (**b** and **e**) in the domesticated dataset as well.

Method of machine learning algorithm comparison: Three machine learning methods, the artificial neural networks (ANNs), the support vector classifier (SVC), and the random forest (RF), are investigated on the confidence evaluation.

The first method, ANNs, is consisted of three layers: the input layer, the hidden layer and the output layer. The parameters, the connection weights between each node, and the thresholds in hidden nodes, are trained by the gradient descent method. Furthermore, in our work, the hidden node number is set as 30.

The second method, SVC, comprises a hyperplane to a high space for the classification. The algorithm is based on the empirical risk minimization to make an intelligent decision. Furthermore, for the nonlinear mapping, the radius basic function is chosen as the kernel for our application.

The last method, RF, is a classical ensemble learning approach, that combines weak classifiers for the decision making. Besides, the random forest, consisted of a series of decision trees. The features are randomly chosen for each tree training. This design can efficiently overcome the problems resulted from the unbalance data. Moreover, the parameter setting for our work is: the tree number 90, and the maximization feature 6.

In addition, for the comparison on the performance of the machine learning method, the cross-validation is applied. We randomly divided the dataset into 4 subsets. Then, we calculated the average performance of the prediction in each subset using the machine learning model training by the remaining 3 subsets. Furthermore, two curves, the ROC curve, and the precision-recall curve, are used to evaluate the forecasting performance for each class (Class 1, Class 2, Class 3, and Class4). The ROC curve can describe the relationship between the true positive ratio (sensitivity) and false positive ratio., especially in balanced data. The second curve, the precision-recall curve, plots the precision and the recall for different thresholds, that is more informative for the imbalanced data evaluation. Furthermore, the index, AUC, is calculated for the ROC curve and precision-recall curve, respectively, to show the forecasting performance for each class. Moreover, the closer to 1 the AUC value is, the better classification for the class is.

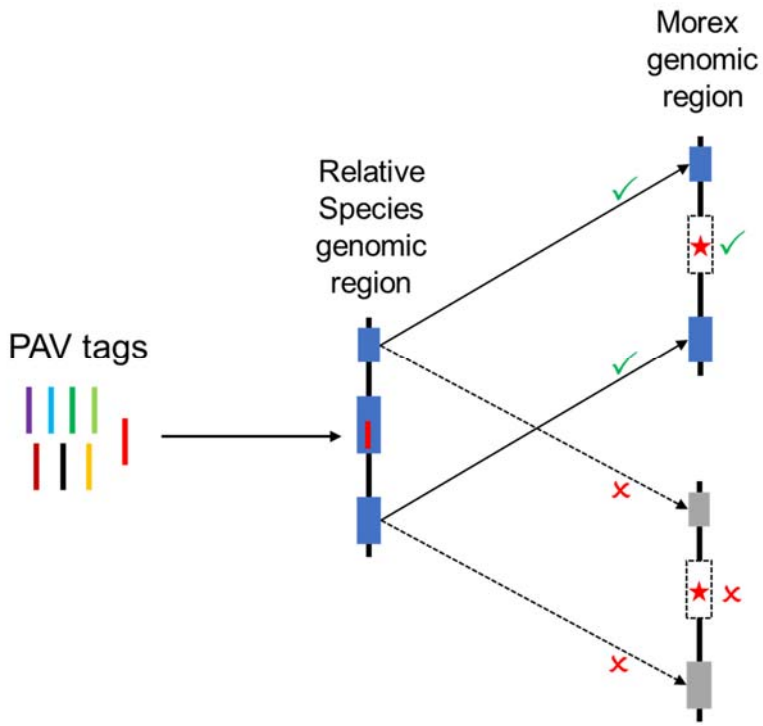


Fig. S6. 5 Validation strategy for PAV-I tags. The red stars indicated the LDM positions of the PAV-I tags.

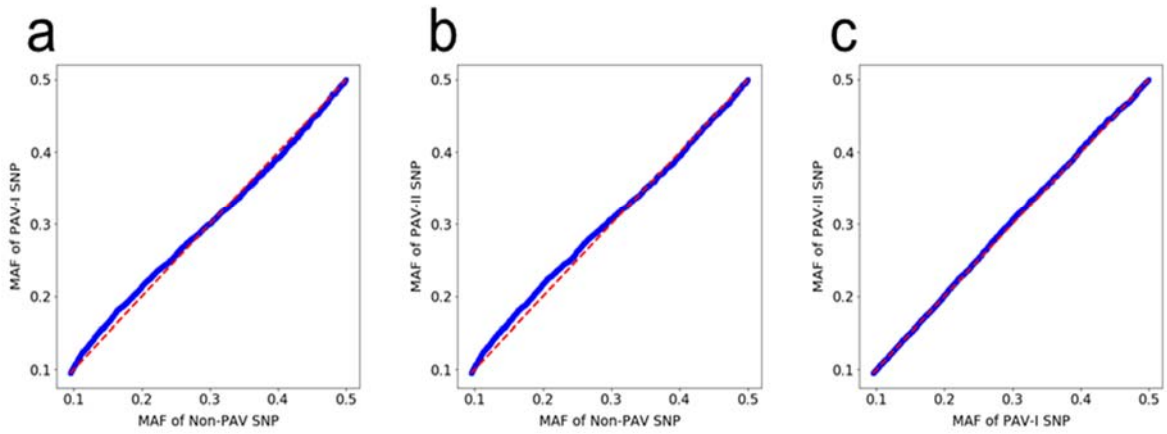


Fig. S6. 6 Q-Q plot of MAF (> 0.095) pairwise comparison for a) Non-PAV SNPs vs PAV-I SNPs; b) Non-PAV SNPs vs PAV-II SNPs; and c) PAV-I SNPs vs PAV-II SNPs

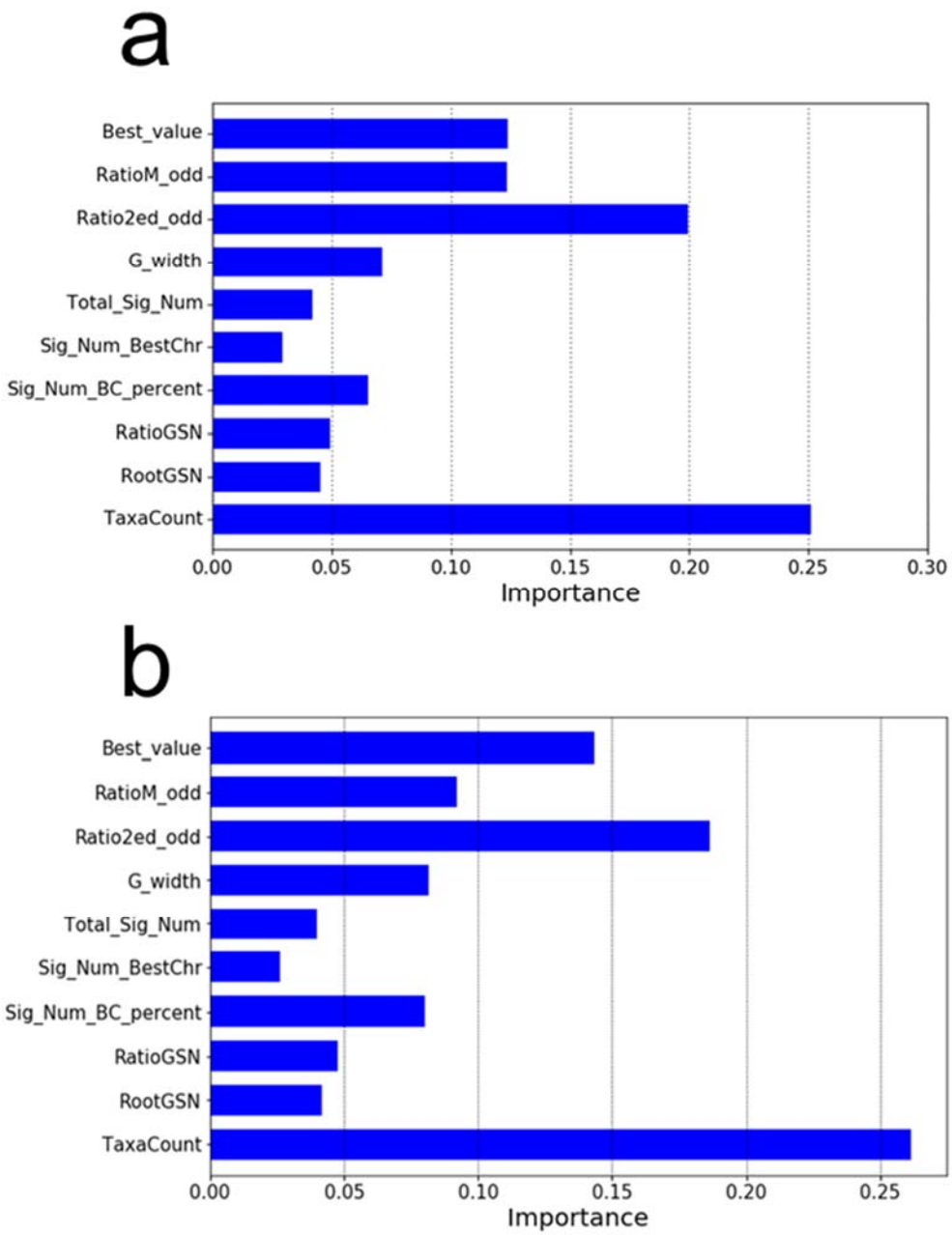


Fig. S6. 7 Importance of features used in random forest model training. (a) data based on the wild barley genotypes; (b) data from the domesticated genotypes.

Supporting tables

Table S3.1 Chromosome assignment and length of linkage groups, and marker density of the maps based on the population of Fleet/AWCS799

Chromosome	Number of markers	Length (cM)	Marker density (cM/Marker)
1H	84	256.3	3.1
2H	105	324.1	3.1
3H	165	356.7	2.2
4H	82	222.0	2.7
5H	134	343.8	2.6
6H	95	199.3	2.1
7H	76	262.5	3.5
Total	741	1964.7	2.7

Table S3. 2 A linkage map of barley based on the population of Fleet/AWCS799

Available at:

https://www.frontiersin.org/articles/10.3389/fpls.2019.01206/full#supplementary-material

Table S3.3 Primer sequences of the maker linked with *Qcrs.caf-6H*

Marker ID	Primer sequences	Product length (bp)	Motif / Number [a]	Tm (°C)
6H_SSR_497772849	5'-GCATTAGTTGTCATAGTAGGTAGCA-3'	242	C / 10	60
	5'-TTCAAGACCACGACCTTGGG-3'			

[a] The difference of motif number between two alleles.

Table S4.1 Primer sequences and results of qPCR validation of RNA-Seq experiments.

Gene ID	Primer 5' -> 3'	Genotype / Mock_vs_FP	log_2FC Cuffdiff	log_2FC qPCR
HORVU1Hr1G092240	Forward : GGAAGCGGCCACTAGCAAAA	NIL1_R	1.79	2.42
	Reverse : ACCAAATCCGCCATCACCAA	NIL1_S	0.85	0.26
		NIL2_R	6.21	6.14
		NIL2_S	0.32	-0.59
		NIL3_R	3.31	3.13
		NIL3_S	4.46	3.80
HORVU1Hr1G092250	Forward : AAACTCGGCGAAGGTGGATT	NIL1_R	1.46	1.76
	Reverse : TCTTCACTGCGATCTCGTCG	NIL1_S	1.89	1.92
		NIL2_R	2.62	2.16
		NIL2_S	2.43	2.96
		NIL3_R	2.87	3.91
		NIL3_S	2.16	2.13
HORVU1Hr1G092300	Forward : GCTAGCTCCACTATGCAGACC	NIL1_R	1.63	2.46
	Reverse : CATCCTCCCATAGCACACGC	NIL1_S	3.75	3.84
		NIL2_R	3.75	4.82
		NIL2_S	3.13	3.70
		NIL3_R	0.95	0.07
		NIL3_S	2.96	3.28

Note: qPCR results for 3 selected DEGs between the mock and inoculated isolines among the three pairs of NILs. The fold-change of qPCR results for each gene was generally in agreement with RNA-seq results.

Table S4.2 to S4.6 are available at:

https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-6011-8#Sec18

Table S4.2 Enriched GO terms associated with DEGs and HEGs. In the comparison column, 'M' = mock-inoculation; 'I'=*Fp*-inoculation; 'R' = resistant isoline; 'S' = susceptible isoline. 'O' column stands for three domains, 'C' = cellular component; 'F' = molecular function; 'P' = biological process. '#list' means the number of term-specific genes from the input list. '#bg' means the number of term-specific genes from the background. FDR <0.05.

Table S4.3 DEGs and SNP-bearing genes within the SNP consensus region across the three NIL pairs. Log2Fold Changes for each of the genes in different comparisons were listed (FDR <= 0.05). 'M' = mock-inoculation; 'I' = *Fp*-inoculation; 'R' = resistant isoline; 'S' = susceptible isoline. Positive values mean that the gene was up-regulated following *Fp*-inoculation; and negative values indicate down-regulated genes. 'inf' means the value of the comparative object is zero.

Table S4.4 Annotation of non-synonymous SNPs in genes within the consensus region. $: "-" means that SNPs were not found in the high confidence (HC) gene. * blank cell means no amino acid change was detected.

Table S4.5 DEGs related to typical resistance mechanisms against *F. graminearum and F. pseudograminearum*. Log2Fold Changes for each gene in different comparisons were listed (FDR <= 0.05). 'M' = mock-inoculation; 'I' = *Fp*-inoculation; 'R' = resistant isoline; 'S' = susceptible isoline. Positive values mean that the gene was up-regulated following *Fp*-inoculation; negative value indicates down-regulated genes, and 'inf' means the value of the comparative object is zero.

Table S4.6 GO annotations of up- (Sheet 1) and down-regulated (Sheet 2) DEGs and background references (Sheet 3) used in GO enrichment analysis.

Table S5.1 Markers based on insertions/deletions (Indels) targeting the genomic interval containing the *Qcrs.cpi-1H* locus

Markers	Physical position	Forward primer (5' -> 3')	Reverse primer (5' -> 3')	Ann. T (°C)
Sgs_5463	546343412	GCGAGAGGGACATGAGGATG	GGAGAAAAAGGAGGACGCGA	60
Sgs_5484	548440405	TGTTGCCAGTTCAGGTGTGT	TCTTTTCCGTCTGCTTTCGC	55
Sgs_5488	548899896	TGCGTTTGTACTGGGATTTGC	GATGTCTCCGCTCCACCATT	55
WMC1E8	548929256	TCATTCGTTGCAGATACACCAC	TCAATGCCCTTGTTTCTGACCT	55
Sgs_5514	551441548	TTCCAGGAAGATCCCATCTG	TGTGGCAAAGCTCACAAAAG	50
Sgs_5515	551526178	ATCGAGCAAGCAAGAAGAGG	GCGTATATTCCCTTGGTGGA	50
Sgs_5508	550802298	TAGGTCAAACTGGGGTGTCG	CACGTCGTGCTCCACTAAGA	55
Sgs_5511	551196802	CTGCTTGGCTCGTTAAGCTC	GCTCCTAATCCTGTTGTCACC	55

Table S5.2 Kompetitive allele specific PCR (KASP) markers designed for the genomic region of *Qcrs.cpi-1H*

Markers	Physical position		Forward primer (5' -> 3') *	Common reverse primer (5' -> 3')	Ann. T (°C)
Kgs_5491.5	549155475	R	GTGATGGGTGAAAACAAG**A**	TACTCCGACTGTAAACGGAA	60
		S	GTGATGGGTGAAAACAAG**G**		
Kgs_5491.8	549186864	R	AACGCACACGCCGCGGCG**A**	ACGGCAAGTCGGCAGCAAAG	60
		S	AACGCACACGCCGCGGCG**G**		
Kgs_5493	549304759	R	CGCCGTGGTCAGCGCGCG**A**	CGGGAGGAAGCGGGATTGGA	60
		S	CGCCGTGGTCAGCGCGCG**C**		
Kgs_5497	549787021	R	AAATGTTTGTTGTCCAGG**G**	TGTGTAGCTGCAACATACTG	65
		S	AAATGTTTGTTGTCCAGG**T**		
Kgs_5498	549810091	R	TCATCAACGCGCGGCAGC**A**	CGGTTGACCAACTCGCCGAC	60
		S	TCATCAACGCGCGGCAGC**G**		

Kgs_5502.5	550253355	R	AGGGTATTCTTCCAGACGG**C**	TCCTTGGGTTGAGCTCTTCG	60
		S	AGGGTATTCTTCCAGACGG**A**		
Kgs_5502.4	550244053	R	GCAGAGCAACGTGAAGGCC**C**	TTCCCGATGGTCACGCCATT	60
		S	GCAGAGCAACGTGAAGGCC**T**		
Kgs_5503	550315683	R	ACCGGGTCCAGATTATTCCAG**C**	CAGCCCGCAAAGTTGAAATGT	60
		S	ACCGGGTCCAGATTATTCCAG**T**		
Kgs_5505.1	550543149	R	CCACATTGTGGAGGTACTC**T**	ACCATAGTATCCGCTTGGCA	60
		S	CCACATTGTGGAGGTACTC**A**		
Kgs_5505.2	550546329	R	CCACATTGTGGAGGTACTC**T**	ACCATAGTATCCGCTTGGCA	65
		S	CCACATTGTGGAGGTACTC**A**		
Kgs_5505.3	550557941	R	GTAGTCTTAAGGACATTAT**A**	TGAAGAAAAAGGTCAATACC	55
		S	GTAGTCTTAAGGACATTAT**T**		
Kgs_5505.4	550557969	R	ATTTGGTTCTCATACCCAA**T**	GTTCTGATGTCCCTGTCTTG	55
		S	ATTTGGTTCTCATACCCAA**C**		
Kgs_5506.1	550659750	R	CAGATGTGACAGAACCAGG**A**	AGATCCCAGCAGAGAGGTGT	65
		S	CAGATGTGACAGAACCAGG**G**		
Kgs_5506.2	550659770	R	AGATGTGACAGAACCAGG**A**	TCGTTGAGGCCTCTCTAACA	60
		S	AGATGTGACAGAACCAGG**G**		

* Forward primers of resistant allele were capped with HEX-tag: 5'-GAAGGTCGGAGTCAACGGATT-3'; forward primers of susceptible allele were capped with FAM-tag: 5'-GAAGGTGACCAAGTTCATGCT-3'.

Table S5.3 High- and low-confidence genes in the genomic region harbouring *Qcrs.cpi-1H*

Predicted gene ID [a]	Physical location on 1H pseudomolecule (bp)	Confidence class [b]	Annotation [c]
HORVU1Hr1G092310	550,318,133 - 550,320,269	HC_G	Glucan endo-1,3-beta-glucosidase 13
HORVU1Hr1G092320	550,323,309 - 550,324,570	LC_TE	Retrotransposon protein, putative, Ty1-copia subclass
HORVU1Hr1G092330	550,383,701 - 550,385,548	HC_G	Sugar transporter protein 7
HORVU1Hr1G092340	550,439,171 - 550,439,691	HC_G	S-type anion channel SLAH2
HORVU1Hr1G092350	550,456,653 - 550,457,250	HC_G	Sulfate transporter 4;2
HORVU1Hr1G092360	550,471,300 - 550,472,320	HC_G	WRKY DNA-binding protein 27
HORVU1Hr1G092370	550,484,748 - 550,485,149	HC_G	Ycf68
HORVU1Hr1G092380	550,486,061 - 550,486,775	LC_U	Chromosome 3B, genomic scaffold, cultivar Chinese Spring
HORVU1Hr1G092390	550,488,755 - 550,490,255	HC_G	Cell wall-associated hydrolase
HORVU1Hr1G092400	550,489,104 - 550,489,278	HC_G	Cell wall-associated hydrolase
HORVU1Hr1G092410	550,492,483 - 550,492,674	LC_u	- [c]
HORVU1Hr1G092420	550,493,104 - 550,493,376	HC_G	30S ribosomal protein S15, chloroplastic
HORVU1Hr1G092430	550,495,135 - 550,495,387	HC_U	-
HORVU1Hr1G092440	550,543,149 - 550,549,150	HC_G	P-loop containing nucleoside triphosphate hydrolases superfamily protein
HORVU1Hr1G092450	550,544,502 - 550,544,868	LC_u	-
HORVU1Hr1G092460	550,557,824 - 550,558,303	LC_u	-
HORVU1Hr1G092470	550,570,184 - 550,570,739	LC_u	-
HORVU1Hr1G092480	550,570,330 - 550,570,682	LC_u	-
HORVU1Hr1G092490	550,625,635 - 550,625,795	LC_u	-
HORVU1Hr1G092500	550,629,626 - 550,630,250	LC_u	-
HORVU1Hr1G092510	550,644,080 - 550,644,250	LC_TE	-
HORVU1Hr1G092520	550,659,293 - 550,660,177	LC_u	-
HORVU1Hr1G092530	550,688,424 - 550,692,028	LC_TE	Gag-pol polyprotein

HORVU1Hr1G092540	550,793,659 - 550,794,979	HC_G	Late embryogenesis abundant hydroxyproline-rich glycoprotein family
HORVU1Hr1G092550	550,797,550 - 550,799,398	HC_G	LRR receptor-like serine/threonine-protein kinase GSO2
HORVU1Hr1G092560	550,797,624 - 550,799,366	LC_u	-
HORVU1Hr1G092570	550,799,948 - 550,801,745	LC_u	-
HORVU1Hr1G092580	550,801,891 - 550,803,181	LC_u	-

[a] The predicted genes and their annotations were retrieved from IPK-Gatersleben blast-website (https://webblast.ipk-gatersleben.de/barley_ibsc/downloads/). Genes in bold hold SNPs between resistant and susceptible alleles (More details in Table S4) [b] The confidence class is provided by the annotation files mentioned in note [a]. *HC_G*: high-confidence gene with predicted function; *LC_u*: low-confidence gene without functional annotation; *HC_TE*: high confidence genes which might be transposable element due to conflicting information; *LC_TE*: genes that are annotated as transposable elements

[c] No definite functional annotation

Table S5.4 Types of SNP variations within the targeted interval of *Qcrs.cpi-1H*

Gene ID	Gene description	Position	R allele	S allele	Variation type
HORVU1Hr1G092310	Glucan endo-1,3-beta-glucosidase 13	550318397	T	C	missense_variant Ala548Thr
		550318461	C	T	synonymous_variant
		550318572	C	T	synonymous_variant
		550318755	A	C	missense_variant Glu428Asp
		550318941	A	G	synonymous_variant
		550318959	G	A	synonymous_variant
		550319001	A	G	synonymous_variant
		550319152	C	G	missense_variant Pro296Arg
		550319153	T	G	missense_variant Pro296Thr
		550319186	G	A	missense_variant Tyr285His
		550319546	C	G	missense_variant Leu165Val
		550319586	C	A	synonymous_variant
		550320014	G	C	missense_variant Val9Leu
HORVU1Hr1G092330	Sugar transporter protein 7	550383847	A	G	synonymous_variant
HORVU1Hr1G092440	P-loop containing nucleoside triphosphate hydrolases superfamily protein	550544005	T	G	splice_region_variant&intron_variant
		550544318	A	G	splice_donor_variant&intron_variant
		550545485	C	T	splice_region_variant&intron_variant
		550546349	T	A	missense_variant Ser110Thr
		550546516	T	G	splice_donor_variant&intron_variant
		550546577	G	A	3_prime_UTR_variant
		550547629	A	G	splice_donor_variant&intron_variant
		550547869	A	G	splice_donor_variant&intron_variant
		550548215	A	G	splice_donor_variant&intron_variant
		550548418	A	C	3_prime_UTR_variant

		550548606	G	T	3_prime_UTR_variant
		550548680	A	G	3_prime_UTR_variant
		550548712	G	T	3_prime_UTR_variant
		550548828	T	C	3_prime_UTR_variant
		550548915	A	C	3_prime_UTR_variant
HORVU1Hr1G092460	-	550557908	T	G	intergenic_region
		550557961	A	T	intergenic_region
		550557989	T	C	intergenic_region
		550558092	A	G	intergenic_region
		550558233	G	C	intergenic_region
		550558242	G	A	intergenic_region
HORVU1Hr1G092520	-	550659770	A	G	intergenic_region
		550660016	A	G	intergenic_region
HORVU1Hr1G092540	Late embryogenesis abundant hydroxyproline-rich glycoprotein family	550793985	C	T	synonymous_variant
		550794042	C	T	synonymous_variant
		550794366	G	A	synonymous_variant
		550794710	C	G	3_prime_UTR_variant
		550794782	A	G	3_prime_UTR_variant
		550794815	T	C	3_prime_UTR_variant
HORVU1Hr1G092550	LRR receptor-like serine/threonine-protein kinase GSO2	550797770	G	C	5_prime_UTR_variant
		550797775	A	T	5_prime_UTR_variant
		550798003	C	T	5_prime_UTR_premature_start_codon_gain_variant
		550798141	A	T	5_prime_UTR_premature_start_codon_gain_variant
		550798255	A	G	5_prime_UTR_premature_start_codon_gain_variant
		550798342	C	G	5_prime_UTR_variant
		550798395	T	C	5_prime_UTR_variant
		550798541	A	C	5_prime_UTR_variant

	550798960	A	C	downstream_gene_variant
	550799048	T	C	downstream_gene_variant
	550799298	A	C	synonymous_variant

Table S6.1 to S6.7 are available at: https://github.com/ShawnGao911101/LDM-RF
File name: Supporting Tables Table S1-S7.xlsx

Table S6.1 European Nucleotide Archive (ENA) IDs of the wild barley genotypes used in this study

Table S6.2 European Nucleotide Archive (ENA) IDs of domesticated barley genotypes used in this study

Table S6.3 Features used for model training in machine learning

Table S6.4 Barley genotypes and associated phenotypic data used in the GWAS analyses.

Table S6.5 Numbers of UAMTs in Classes 1 to 4

Table S6.6 Validation results of PAVI-tags from the domesticated barley genotypes. Note: Column 'Distance' represents the distance between LDM position and Physical position on Morex.

Table S6.7 Validation results of PAVI-tags from the wild barley genotypes.

Note: Column 'Distance' represents the distance between LDM position and Physical position on Morex.

Bibliography

1001 Genomes Consortium (2016) 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. Cell 166:481-491

Akinsanmi OA, Mitter V, Simpfendorfer S, Backhouse D, Chakraborty S (2004) Identity and pathogenicity of *Fusarium* spp. isolated from wheat fields in Queensland and northern New South Wales. Crop and Pasture Science 55:97-107

Allard RW (1999) History of plant population genetics. Annual Review of Genetics 33:1-27

Aoki T, O'Donnell K (1999) Morphological and molecular characterization of Fusarium pseudograminearum sp. nov., formerly recognized as the Group 1 population of *F. graminearum*. Mycologia:597-609

Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361(6403). https://doi.org/10.1126/science.aar7191

Ariyadasa R, Mascher M, Nussbaumer T, Schulte D, Frenkel Z, Poursarebani N, Zhou R, Steuernagel B, Gundlach H, Taudien S (2014) A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. Plant Physiology 164:412-423

Asins M (2002) Present and future of quantitative trait locus analysis in plant breeding. Plant Breeding 121:281-291

Backhouse D, Burgess L (2002) Climatic analysis of the distribution of *Fusarium graminearum, F. pseudograminearum and F. culmorum* on cereals in Australia. Australasian Plant Pathology 31:321-327

Badr A, Rabey HE, Effgen S, Ibrahim H, Pozzi C, Rohde W, Salamini F (2000) On the origin and domestication history of barley (*Hordeum vulgare*). Molecular Biology and Evolution 17:499-510

Bai Z, Liu C (2015) Histological Evidence for Different Spread of Fusarium crown rot in Barley Genotypes with Different Heights. Journal of Phytopathology 163:91-97

Barrero JM, Cavanagh C, Verbyla KL, Tibbits JF, Verbyla AP, Huang BE, Rosewarne GM, Stephen S, Wang P, Whan A (2015) Transcriptomic analysis of wheat near-isogenic lines identifies PM19-A1 and A2 as candidates for a major dormancy QTL. Genome Biology 16:93

Bates D, Sarkar D, Bates MD, Matrix L (2007) The lme4 package. R package version 2:74

Beavis WD (1998) QTL analyses: power, precision, and accuracy. Molecular Dissection of Complex Traits 1998:145-162

Beddis A, Burgess L (1992) The influence of plant water stress on infection and colonization of wheat seedlings by *Fusarium graminearum* Group 1. Phytopathology 82:78-83

Beffa RS, Neuhaus J-M, Meins F (1993) Physiological compensation in antisense transformants: specific induction of an" ersatz" glucan endo-1, 3-beta-glucosidase in plants infected with necrotizing viruses. Proceedings of the National Academy of Sciences 90:8792-8796

Bentley A, Tunali B, Nicol J, Burgess L, Summerell B A survey of *Fusarium* species associated with wheat and grass stem bases in northern Turkey. SYDOWIA-HORN- 58

Birkenbihl RP, Kracher B, Roccaro M, Somssich IE (2017) Induced genome-wide binding of three *Arabidopsis* WRKY transcription factors during early MAMP-triggered immunity. The Plant Cell 29:20-38

Blencowe BJ, Ahmad S, Lee LJ (2009) Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. Genes & Development 23:1379-1386

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Brkljacic J, Grotewold E, Scholl R, Mockler T, Garvin DF, Vain P, Brutnell T, Sibout R, Bevan M, Budak H (2011) *Brachypodium* as a model for the grasses: today and the future. Plant Physiology 157:3-13

Burgess L (2005) Intermediate hosts and the management of crown rot and head blight. Annual Report of GRDC strategic initiative on crown rot, common root rot and *Fusarium* head blight Grains Research and Development Corporation, Kingston, Australia:34-36

Burgess L (2014) 2011 McAlpine Memorial Lecture-A Love Affair with *Fusarium*. Australasian Plant Pathology 43:359-368

Burgess L, Backhouse D, Swan L, Esdaile R (1996) Control of *Fusarium* crown rot of wheat by late stubble burning and rotation with sorghum. Australasian Plant Pathology 25:229-233

Burgess L, Bryden W (2012) Fusarium: a ubiquitous fungus of global significance. Microbiology Australia 33:22-25

Burgess L, Griffin D (1968) The recovery of Gibberella zeae from wheat straws. Animal Production Science 8:364-370

Burgess L, Wearing A, Toussoun T (1975) Surveys of *Fusaria* associated with crown rot of wheat in eastern Australia. Crop and Pasture Science 26:791-799

Burgess LW, Backhouse D, Summerell BA, Swan LJ (2001) Crown rot of wheat. *Fusarium*: Paul E Nelson Memorial Symposium APS Press, St Paul, Minnesota, pp 1-294

Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the National Academy of Sciences 110:8057-8062

Chakraborty S, Liu C, Mitter V, Scott J, Akinsanmi O, Ali S, Dill-Macky R, Nicol J, Backhouse D, Simpfendorfer S (2006) Pathogen population structure and epidemiology are keys to wheat crown rot and *Fusarium* head blight management. Australasian Plant Pathology 35:643-655

Chakraborty S, Obanor F, Westecott R, Abeywickrama K (2010) Wheat crown rot pathogens *Fusarium graminearum and F. pseudograminearum* lack specialization. Phytopathology 100:1057-1065

Chen G, Habib A, Wei Y, Zheng Y-L, Shabala S, Zhou M, Liu C (2015) Enhancing *Fusarium* crown rot resistance by pyramiding large-effect QTL in barley. Molecular Breeding 35:1-8

Chen G, Li H, Zheng Z, Wei Y, Zheng Y, McIntyre C, Zhou M, Liu C (2012) Characterization of a QTL affecting spike morphology on the long arm of chromosome 3H in barley (*Hordeum vulgare L.*) based on near isogenic lines and a NIL-derived population. Theoretical and Applied Genetics 125:1385-1392

Chen G, Liu Y, Ma J, Zheng Z, Wei Y, McIntyre CL, Zheng Y-L, Liu C (2013a) A Novel and Major Quantitative Trait Locus for Fusarium Crown Rot Resistance in a Genotype of Wild Barley (*Hordeum spontaneum L.*). PLoS One 8:e58040

Chen G, Liu Y, Wei Y, McIntyre C, Zhou M, Zheng Y-L, Liu C (2013b) Major QTL for *Fusarium* crown rot resistance in a barley landrace. Theoretical and Applied Genetics 126:2511-2520

Chen G, Yan W, Liu Y, Wei Y, Zhou M, Zheng Y-L, Manners JM, Liu C (2014) The non-gibberellic acid-responsive semi-dwarfing gene uzu affects *Fusarium* crown rot resistance in barley. BMC Plant Biology 14:22

Cheung M-Y, Li X, Miao R, Fong Y-H, Li K-P, Yung Y-L, Yu M-H, Wong K-B, Chen Z, Lam H-M (2016) ATP binding by the P-loop NTPase OsYchF1 (an unconventional G protein) contributes to biotic but not abiotic stress responses. Proceedings of the National Academy of Sciences 113:2648-2653

Cheung M, Xue Y, Zhou L, Li M, Sun SS, Lam H (2010) An ancient P-loop GTPase in rice is regulated by a higher plant-specific regulatory protein. Journal of Biological Chemistry 285:37359-37369

Cheung MY, Li MW, Yung YL, Wen CQ, Lam HM (2013) The unconventional P-loop NTPase OsYchF1 and its regulator OsGAP1 play opposite roles in salinity stress tolerance. Plant Cell and Environment 36:2008-2020

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6:80-92

Clear R, Patrick S, Gaba D, Roscoe M, Demeke T, Pouleur S, Couture L, Ward T, O'Donnell K, Turkington T (2006) Trichothecene and zearalenone production, in culture, by isolates of *Fusarium pseudograminearum* from western Canada. Canadian Journal of Plant Pathology 28:131-136

Cohen J (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:37-46

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676

Cook RJ (1980) *Fusarium* foot rot of wheat and its control in the Pacific Northwest. Plant Dis 64:1061-1066

Cook RJ (2001) Management of wheat and barley root diseases in modern farming systems. Australasian Plant Pathology 30:119-126

Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC bioinformatics 11:485-485

Cuthbert PA, Somers DJ, Thomas J, Cloutier S, Brulé-Babel A (2006) Fine mapping *Fhb1*, a major gene controlling *Fusarium* head blight resistance in bread wheat (*Triticum aestivum L.*). Theoretical and Applied Genetics 112:1465-1472

Dai F, Wang X, Zhang XQ, Chen Z, Nevo E, Jin G, Wu D, Li C, Zhang G (2018) Assembly and analysis of a qingke reference genome demonstrate its close genetic relation to modern cultivated barley. Plant Biotechnology Journal 16:760-770

Deorowicz S, Kokot M, Grabowski S, Debudaj-Grabysz A (2015) KMC 2: fast and resource-frugal k-mer counting. Bioinformatics 31:1569-1576

Desmond OJ, Edgar CI, Manners JM, Maclean DJ, Schenk PM, Kazan K (2005) Methyl jasmonate induced gene expression in wheat delays symptom development by the crown rot pathogen *Fusarium pseudograminearum*. Physiological and Molecular Plant Pathology 67:171-179

Desmond OJ, Manners JM, Stephens AE, Maclean DJ, Schenk PM, Gardiner DM, Munn AL, Kazan K (2008) The *Fusarium* mycotoxin deoxynivalenol elicits hydrogen peroxide production, programmed cell death and defence responses in wheat. Molecular Plant Pathology 9:435-445

Diaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012) Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). PLoS One 7:e33234

Dodman R, Wildermuth G (1989) The effect of stubble retention and tillage practices in wheat and barley on crown rot caused by *Fusarium graminearum*. Plant Protection Quarterly 1:98-99

Dolatabadian A, Bayer PE, Tirnaz S, Hurgobin B, Edwards D, Batley J (2019) Characterisation of disease resistance genes in the Brassica napus pangenome reveals significant structural variation. Plant biotechnology journal

Draper J, Mur LA, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge AP (2001) Brachypodium distachyon. A new model system for functional genomics in grasses. Plant Physiology 127:1539-1555

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Research 38:W64-W70

Ellis RP, Forster BP, Robinson D, Handley L, Gordon DC, Russell JR, Powell W (2000) Wild barley: a source of genes for crop improvement in the 21st century? Journal of Experimental Botany 51:9-17

Emrich SJ, Barbazuk WB, Li L, Schnable PS (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Research 17:69-73

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends in Plant Science 5:199-206

Faghani E, Gharechahi J, Komatsu S, Mirzaei M, Khavarinejad RA, Najafi F, Farsad LK, Salekdeh GH (2015) Comparative physiology and proteomic analysis of two wheat genotypes contrasting in drought tolerance. Journal of Proteomics 114:1-15

Felton W, Marcellos H, Martin R, Alston C, Backhouse D, Burgess L, Herridge D (1998) Chickpea in wheat-based cropping systems of northern New South Wales. II. Influence on biomass, grain yield, and crown rot in the following wheat crop. Australian Journal of Agricultural Research

Francis RG, Burgess L (1977) Characteristics of two populations of Fusarium roseum 'Graminearum'in Eastern Australia. Transactions of the British Mycological Society 68:421-427

Gabur I, Chawla HS, Snowdon RJ, Parkin IA (2019) Connecting genome structural variation with complex traits in crop plants. Theoretical and Applied Genetics 132:733-750

Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Burzynski-Chang EA, Fish TL, Stromberg KA, Sacks GL (2019a) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nature Genetics 51:1044-1051

Gao S, Zheng Z, Hu H, Shi H, Ma J, Liu Y, Wei Y, Zheng Y, Zhou M, Liu C (2019b) A novel QTL conferring *Fusarium* crown rot resistance located on chromosome arm 6HL in barley. BioRxiv:537605

Gao S, Zheng Z, Powell J, Habib A, Stiller J, Zhou M, Liu C (2019c) Validation and delineation of a locus conferring *Fusarium* crown rot resistance on 1HL in barley by analysing transcriptomes from multiple pairs of near isogenic lines. BMC Genomics 20:650

Gatti M, Choulet F, Macadré C, Guérard F, Seng J-M, Langin T, Dufresne M (2018) Identification, molecular cloning and functional characterization of a wheat UDP-glucosyltransferase involved in resistance to *Fusarium* Head Blight and to mycotoxin accumulation. Frontiers in Plant Science 9:1853

Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CKK, Severn-Ellis A, McCombie WR, Parkin IA (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. Nature Communications 7:13390

Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin AC, Schackwitz W, Tyler L (2017) Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. Nature Communications 8:2184

Grewal HS, Graham RD, Rengel Z (1996) Genotypic variation in zinc efficiency and resistance to crown rot disease (*Fusarium graminearum* Schw. Group 1) in wheat. Plant and Soil 186:219-226

Habib A, Powell JJ, Stiller J, Liu M, Shabala S, Zhou M, Gardiner DM, Liu C (2018) A multiple near isogenic line (multi-NIL) RNA-seq approach to identify candidate genes underpinning QTL. Theoretical and Applied Genetics 131:613-624

Habib A, Shabala S, Shabala L, Zhou M, Liu C (2016) Near-isogenic lines developed for a major QTL on chromosome arm 4HL conferring *Fusarium* crown rot resistance in barley. Euphytica 209:555-563

Hameed M, Rana R, Ali Z (2012) Identification and characterization of a novel Iraqi isolate of *Fusarium pseudograminearum* causing crown rot in wheat. Genetics and Molecular Research 11:1341-1348

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K (2014) Insights into the maize pan-genome and pan-transcriptome. The Plant Cell 26:121-135

Hogg A, Johnston R, Johnston J, Klouser L, Kephart K, Dyer A (2010) Monitoring *Fusarium* crown rot populations in spring wheat residues using quantitative real-time polymerase chain reaction. Phytopathology 100:49-57

Høj PB, Hartman DJ, Morrice NA, Doan DN, Fincher GB (1989) Purification of (1→ 3)-β-glucan endohydrolase isoenzyme II from germinated barley and determination of its primary structure from a cDNA clone. Plant Molecular Biology 13:31-42

Huang X, Kurata N, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490:497

Huang Y, Li L, Smith KP, Muehlbauer GJ (2016) Differential transcriptomic responses to *Fusarium graminearum* infection in two barley quantitative trait loci associated with Fusarium head blight resistance. BMC Genomics 17:387

International Barley Genome Sequencing Consortium a (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711-716

International Brachypodium Initiative a (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763-768

Jiang Y, Habib A, Zheng Z, Zhou M, Wei Y, Zheng Y-L, Liu C (2019) Development of tightly linked markers and identification of candidate genes for *Fusarium* crown rot resistance in barley by exploiting a near-isogenic line-derived population. Theoretical and Applied Genetics 132:217-225

Jin M, Liu H, He C, Fu J, Xiao Y, Wang Y, Xie W, Wang G, Yan J (2016) Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. Scientific Reports 6:18936

Juliana P, Poland J, Huerta-Espino J, Shrestha S, Crossa J, Crespo-Herrera L, Toledo FH, Govindan V, Mondal S, Kumar U (2019) Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. Nature Genetics: 51(10): 1530-1539

Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database 2011

Kirkegaard J, Simpfendorfer S, Holland J, Bambach R, Moore K, Rebetzke G (2004) Effect of previous crops on crown rot and yield of durum and bread wheat in northern NSW. Crop and Pasture Science 55:321-334

Klein T, Burgess L, Ellison F (1989) The incidence of crown rot in wheat, barley and triticale when sown on two dates. Australian Journal of Experimental Agriculture 29:559-563

Kølster P, Munk L, Stølen O, Løhde J (1986) Near-Isogenic Barley Lines with Genes for Resistance to Powdery Mildew 1. Crop Science 26:903-907

Krattinger SG, Keller B (2016) Molecular genetics and evolution of disease resistance in cereals. New Phytologist 212:320-332

Künzel G, Korzun L, Meister A (2000) Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 154:397-412

Li G, Zhou J, Jia H, Gao Z, Fan M, Luo Y, Zhao P, Xue S, Li N, Yuan Y (2019) Mutation of a histidine-rich calcium-binding-protein gene in wheat confers resistance to *Fusarium* head blight. Nature Genetics 51:1106-1112

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997

Li H, He X, Ding S, Yuan H, Chen L (2016) First Report of *Fusarium culmorum* Causing Crown Rot of Wheat in China. Plant Disease 100:2532-2532

Li H, Yuan H, Fu B, Xing X, Sun B, Tang W (2012) First report of F*usarium pseudograminearum* causing crown rot of wheat in Henan, China. Plant Disease 96:1065-1065

Li HB, Xie GQ, Ma J, Liu GR, Wen SM, Ban T, Chakraborty S, Liu CJ (2010) Genetic relationships between resistances to *Fusarium* head blight and crown rot in bread wheat (Triticum aestivum L.). Theoretical and Applied Genetics 121:941-950

Li HB, Zhou M, Liu CJ (2009) A major QTL conferring crown rot resistance in barley and its association with plant height. Theoretical and Applied Genetics 118:903-910

Li X, Liu C, Chakraborty S, Manners JM, Kazan K (2008) A simple method for the assessment of crown rot disease severity in wheat seedlings inoculated with *Fusarium pseudograminearum*. Journal of Phytopathology 156:751-754

Li X, Michlmayr H, Schweiger W, Malachova A, Shin S, Huang Y, Dong Y, Wiesenberger G, McCormick S, Lemmens M (2017) A barley UDP-glucosyltransferase inactivates nivalenol and provides *Fusarium* Head Blight resistance in transgenic wheat. Journal of Experimental Botany 68:2187-2197

Li Y-h, Zhou G, Ma J, Jiang W, Jin L-g, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nature Biotechnology 32:1045

Liddell C, Burgess L (1988) Wax partitioned soil columns to study the influence of soil moisture potential on the infection of wheat by *Fusarium graminearum* Group 1. Phytopathology 78:185-189

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133:523-536

Liu C, Ogbonnaya FC (2015) Resistance to Fusarium crown rot in wheat and barley: a review. Plant Breeding 134:365-372

Liu H, Zwer P, Wang H, Liu C, Lu Z, Wang Y, Yan G (2016a) A fast generation cycling system for oat and triticale breeding. Plant Breeding 135:574-579

Liu M, Li Y, Ma Y, Zhao Q, Stiller J, Feng Q, Tian Q, Liu D, Han B, Liu C (2020) The draft genome of a wild barley genotype reveals its enrichment in genes related to biotic and abiotic stresses compared to cultivated barley. Plant Biotechnology Journal 18:443-456

Liu M, Stiller J, Holušová K, Vrána J, Liu D, Doležel J, Liu C (2016b) Chromosome-specific sequencing reveals an extensive dispensable genome component in wheat. Scientific Reports 6:36398

Liu RH, Meng JL (2003) MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data. Hereditas 25:317-321

Liu X, Liu C (2016) Effects of drought-stress on *Fusarium* crown rot development in Barley. PLoS One 11:e0167304

Liu Y, Ma J, Yan W, Yan G, Zhou M, Wei Y, Zheng Y, Liu C (2012a) Different tolerance in bread wheat, durum wheat and barley to *Fusarium* crown rot disease caused by *Fusarium pseudograminearum*. Journal of Phytopathology 160:412-417

Liu Y, Zheng YL, Wei Y, Zhou M, Liu C (2012b) Genotypic differences to crown rot caused by *Fusarium pseudograminearum* in barley (*Hordeum vulgare L.*). Plant Breeding 131:728-732

Liu YX, Yang XM, Ma J, Wei YM, Zheng YL, Ma HX, Yao JB, Yan GJ, Wang YG, Manners JM (2010) Plant height affects *Fusarium* crown rot severity in wheat. Phytopathology 100:1276-1281

Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang T, Li Y, Li Y, Semagn K, Zhang X (2015) High-resolution genetic mapping of maize pan-genome sequence anchors. Nature Communications 6:6914

Ma J, Li HB, Zhang CY, Yang XM, Liu YX, Yan GJ, Liu CJ (2010) Identification and validation of a major QTL conferring crown rot resistance in hexaploid wheat. Theoretical and Applied Genetics 120:1119-1128

Ma J, Stiller J, Zhao Q, Feng Q, Cavanagh C, Wang P, Gardiner D, Choulet F, Feuillet C, Zheng Y-L (2014) Transcriptome and Allele Specificity Associated with a 3BL Locus for *Fusarium* Crown Rot Resistance in Bread Wheat. PLos One 9:1-11

Ma J, Yan GJ, Liu CJ (2012) Development of near-isogenic lines for a major QTL on 3BL conferring *Fusarium* crown rot resistance in hexaploid wheat. Euphytica 183:147-152

Ma Y, Liu M, Stiller J, Liu C (2019) A pan-transcriptome analysis shows that disease resistance genes have undergone more selection pressure during barley domestication. BMC Genomics 20:1-11

Mackill D, Bonman J (1992) Inheritance of blast resistance in near-isogenic lines of rice. Phytopathology 82:746-749

Marone D, Russo M, Laidò G, De Leonardis A, Mastrangelo A (2013) Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. International Journal of Molecular Sciences 14:7302-7326

Martin A, Bovill W, Percy C, Herde D, Fletcher S, Kelly A, Neate S, Sutherland M (2015) Markers for seedling and adult plant crown rot resistance in four partially resistant bread wheat sources. Theoretical and Applied Genetics 128:377-385

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J (2017) A chromosome conformation capture ordered sequence of the barley genome. Nature 544:427-433

McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiology 159:1295-1308

McKnight T, Hart J (1966) Some field observations on crown rot disease of wheat caused by Fusarium graminearum. Queensland Journal of Agricultural and Animal Sciences 23:373-378

Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. The Crop Journal 3:269-283

Miles C, Wayne M (2008) Quantitative trait locus (QTL) analysis. Nature Education, 1 (1).

Milne RJ, Dibley KE, Schnippenkoetter W, Mascher M, Lui AC, Wang L, Lo C, Ashton AR, Ryan PR, Lagudah ES (2019) The Wheat Lr67 Gene from the Sugar Transport Protein 13 Family Confers Multipathogen Resistance in Barley. Plant Physiology 179:1285-1297

Milner SG, Jost M, Taketa S, Mazón ER, Himmelbach A, Oppermann M, Weise S, Knüpffer H, Basterrechea M, König P (2019) Genebank genomics highlights the diversity of a global barley collection. Nature Genetics 51:319-326

Mitter V, Zhang MC, Liu CJ, Ghosh R, Ghosh M, Chakraborty S (2006) A high-throughput glasshouse bioassay to detect crown rot resistance in wheat germplasm. Plant Pathology 55:433-441

Monat C, Padmarasu S, Lux T, Wicker T, Gundlach H, Himmelbach A, Ens J, Li C, Muehlbauer GJ, Schulman AH (2019a) TRITEX: chromosome-scale sequence assembly of *Triticeae* genomes with open-source tools. BioRxiv:631648

Monat C, Schreiber M, Stein N, Mascher M (2019b) Prospects of pan-genomics in barley. Theoretical and Applied Genetics 132:785-796

Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan CKK, Visendi P, Lai K, Doležel J, Batley J (2017) The pangenome of hexaploid bread wheat. The Plant Journal 90:1007-1013

Moore JW, Herrera-Foessel S, Lan C, Schnippenkoetter W, Ayliffe M, Huerta-Espino J, Lillemo M, Viccars L, Milne R, Periyannan S (2015) A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. Nature Genetics 47:1494–1498

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5:621-628

Mudge AM, Dill-Macky R, Dong Y, Gardiner DM, White RG, Manners JM (2006) A role for the mycotoxin deoxynivalenol in stem colonisation during crown rot disease of wheat caused by *Fusarium graminearum* and *Fusarium pseudograminearum*. Physiological and Molecular Plant Pathology 69:73-85

Murray G, Brennan J (2010) Estimating disease losses to the Australian barley industry. Australasian Plant Pathology 39:85-96

Murray GM, Brennan JP (2009) Estimating disease losses to the Australian wheat industry. Australasian Plant Pathology 38:558-570

Oliveros J (2018) An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/

Ott A, Liu S, Schnable JC, Yeh C-TE, Wang K-S, Schnable PS (2017) tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Research 45:e178-e178

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124:735-742

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726

Paulitz TC, Smiley RW, Cook RJ (2002) Insights into the prevalence and management of soilborne cereal pathogens under direct seeding in the Pacific Northwest, USA. Canadian Journal of Plant Pathology 24:416-428

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12:2825-2830

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253

Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter 15:8-15

Powell JJ, Carere J, Fitzgerald TL, Stiller J, Covarelli L, Xu Q, Gubler F, Colgrave ML, Gardiner DM, Manners JM (2017a) The *Fusarium* crown rot pathogen *Fusarium pseudograminearum* triggers a suite of transcriptional and metabolic changes in bread wheat (Triticum aestivum L.). Annals of Botany 119:853-867

Powell JJ, Carere J, Sablok G, Fitzgerald TL, Stiller J, Colgrave ML, Gardiner DM, Manners JM, Vogel JP, Henry RJ (2017b) Transcriptome analysis of *Brachypodium* during fungal pathogen infection reveals both shared and distinct defense responses with wheat. Scientific Reports 7:17212

Pumphrey MO, Bernardo R, Anderson JA (2007a) Validating the Fhb1 QTL for Fusarium head blight resistance in near-isogenic wheat lines developed from breeding populations. Crop Science 47:200-206

Pumphrey MO, Bernardo R, Anderson JA (2007b) Validating the QTL for *Fusarium* head blight resistance in near-isogenic wheat lines developed from breeding populations. Crop Science 47:200-206

Purss G (1966) Studies of varietal resistance to crown rot of wheat caused by *Fusarium graminearum* Schw. Queensland Journal of Agricultural and Animal Sciences 23:475-498

Purss G (1971) Pathogenic specialization in *Fusarium graminearum*. Crop and Pasture Science 22:553-561

Rawat N, Pumphrey MO, Liu S, Zhang X, Tiwari VK, Ando K, Trick HN, Bockus WW, Akhunov E, Anderson JA (2016) Wheat *Fhb1* encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to *Fusarium* head blight. Nature Genetics 48:1576

Revelle WR (2017) psych: Procedures for personality and psychological research. https://cran.r-project.org/web/packages/psych/

Ribaut J-M, Hoisington D (1998) Marker-assisted selection: new tools and strategies. Trends in Plant Science 3:236-239

Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner H-Y, Hunt MD (1996) Systemic Acquired Resistance. The Plant Cell 8:1809-1819

Samans B, Chalhoub B, Snowdon RJ (2017) Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species Brassica napus. The Plant Genome https://doi.org/10.3835/plantgenome2017.02.0013

SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765-768

Saxena RK, Edwards D, Varshney RK (2014) Structural variations in plant genomes. Briefings in Functional Genomics 13:296-307

Schiessl S-V, Katche E, Ihien E, Chawla HS, Mason AS (2019) The role of genomic structural variation in the genetic improvement of polyploid crops. The Crop Journal 7:127-140

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112-1115

Schreiber M, Stein N, Mascher M (2018) Genomic approaches for studying crop evolution. Genome Biology 19:140

Schulte D, Close TJ, Graner A, Langridge P, Matsumoto T, Muehlbauer G, Sato K, Schulman AH, Waugh R, Wise RP (2009) The international barley sequencing consortium—at the threshold of efficient access to the barley genome. Plant Physiology 149:142-147

Schweiger W, Boddu J, Shin S, Poppenberger B, Berthiller F, Lemmens M, Muehlbauer GJ, Adam G (2010) Validation of a candidate deoxynivalenol-inactivating UDP-glucosyltransferase from barley by heterologous expression in yeast. Molecular Plant-microbe Interactions 23:977-986

Schweiger W, Pasquet J-C, Nussbaumer T, Paris MPK, Wiesenberger G, Macadré C, Ametz C, Berthiller F, Lemmens M, Saindrenan P (2013) Functional characterization of two clusters of *Brachypodium distachyon* UDP-glycosyltransferases encoding putative deoxynivalenol detoxification genes. Molecular Plant-Microbe Interactions 26:781-792

Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH (2017) DNA sequencing at 40: past, present and future. Nature 550:345-353

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nature Biotechnology 26:1135-1145

Shendure J, Mitra RD, Varma C, Church GM (2004) Advanced sequencing technologies: methods and goals. Nature Reviews Genetics 5:335-344

Shin S-J, Ahn H, Jung I, Rhee S, Kim S, Kwon H-B (2016) Novel drought-responsive regulatory coding and non-coding transcripts from *Oryza Sativa L*. Genes & Genomics 38:949-960

Simpfendorfer S, Brettell R, Nicol J (2012) Inter-row sowing reduces crown rot in winter cereals. First International Crown Rot Workshop for wheat improvement' Narrabri Organising Committee of the 1st International Crown Rot Workshop

Singh DP, Backhouse D, Kristiansen P (2009) Interactions of temperature and water potential in displacement of *Fusarium pseudograminearum* from cereal residues by fungal antagonists. Biological Control 48:188-195

Smiley RW (2009) Root-lesion nematodes reduce yield of intolerant wheat and barley. Agronomy Journal 101:1322-1335

Smiley RW, Gourlie JA, Easley SA, Patterson L-M (2005a) Pathogenicity of fungi associated with the wheat crown rot complex in Oregon and Washington. Plant Disease 89:949-957

Smiley RW, Gourlie JA, Easley SA, Patterson L-M, Whittaker RG (2005b) Crop damage estimates for crown rot of wheat and barley in the Pacific Northwest. Plant Disease 89:595-604

Smiley RW, Patterson L-M (1996) Pathogenic fungi associated with *Fusarium* foot rot of winter wheat in the semiarid Pacific Northwest. Plant Disease 80:944-949

Song Y, Jing S, Yu D (2009) Overexpression of the stress-induced OsWRKY08 improves osmotic stress tolerance in *Arabidopsis*. Chinese Science Bulletin 54:4671-4678

Sparrow DH, Graham RD (1988) Susceptibility of zinc-deficient wheat plants to colonization by *Fusarium graminearum* Schw. Group 1. Plant and Soil 112:261-266

Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genetics 5:e1000734

Stark R, Grzelak M, Hadfield J (2019) RNA sequencing: the teenage years. Nature Reviews Genetics 20:631-656

Su Y, Wang Z, Liu F, Li Z, Peng Q, Guo J, Xu L, Que Y (2016) Isolation and characterization of ScGluD2, a new sugarcane beta-1, 3-glucanase D family gene induced by *Sporisorium scitamineum*, ABA, H2O2, NaCl, and CdCl2 stresses. Frontiers in Plant Science 7:1348

Su Z, Bernardo A, Tian B, Chen H, Wang S, Ma H, Cai S, Liu D, Zhang D, Li T (2019a) A deletion mutation in TaHRC confers *Fhb1* resistance to *Fusarium* head blight in wheat. Nature Genetics 52:1099–1105

Su Z, Bernardo A, Tian B, Chen H, Wang S, Ma H, Cai S, Liu D, Zhang D, Li T (2019b) A deletion mutation in TaHRC confers *Fhb1* resistance to *Fusarium* head blight in wheat. Nature Genetics 51: 1099-1105

Summerell B, Burgess L, Klein T (1989) The impact of stubble management on the incidence of crown rot of wheat. Animal Production Science 29:91-98

Sun C, Hu Z, Zheng T, Lu K, Zhao Y, Wang W, Shi J, Wang C, Lu J, Zhang D (2016) RPAN: rice pan-genome browser for ~3000 rice genomes. Nucleic Acids Research 45:597-605

Swan L, Backhouse D, Burgess L (2000) Surface soil moisture and stubble management practice effects on the progress of infection of wheat by *Fusarium pseudograminearum*. Animal Production Science 40:693-698

Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, Su Z, Pan Y, Liu D, Lipka AE (2016) GAPIT version 2: an enhanced integrated tool for genomic association and prediction. The Plant Genome https://doi.org/10.3835/plantgenome2015.11.0120

R Core Team (2013) R: A language and environment for statistical computing.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences 102:13950-13955

Thongbai P, Hannam RJ, Graham RD, Webb MJ (1993) Interaction between zinc nutritional status of cereals and *Rhizoctonia* root rot severity. Plant and Soil 153:207-214

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic acids research 45:W122-W129

Tombuloglu G, Tombuloglu H, Sakcali MS, Unver T (2015) High-throughput transcriptome analysis of barley (*Hordeum vulgare*) exposed to excessive boron. Gene 557:71-81

Tottman D, Makepeace R, Broad H (1979) An explanation of the decimal code for the growth stages of cereals, with illustrations. Annals of Applied Biology 93:221-234

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7:562-578

Tuinstra M, Ejeta G, Goldsbrough P (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theoretical and Applied Genetics 95:1005-1011

Tunali B, Nicol JM, Hodson D, Uckun Z, Büyük O, Erdurmuş D, Hekimhan H, Aktaş H, Akbudak MA, Bağci SA (2008) Root and crown rot fungi associated with spring, facultative, and winter wheat in Turkey. Plant Disease 92:1299-1306

Van Eeuwijk Fv, Mesterhazy A, Kling CI, Ruckenbauer P, Saur L, Buerstmayr H, Lemmens M, Keizer L, Maurin N, Snijders C (1995) Assessing non-specificity of resistance in wheat to head blight caused by inoculation with European strains of *Fusarium culmorum*, *F. graminearum* and *F. nivale* using a multiplicative model for interaction. Theoretical and Applied Genetics 90:221-228

Verrell AG, Simpfendorfer S, Moore KJ (2017) Effect of row placement, stubble management and ground engaging tool on crown rot and grain yield in a no-till continuous wheat sequence. Soil and Tillage Research 165:16-22

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93:77-78

Wallwork H, Butt M, Cheong J, Williams K (2004) Resistance to crown rot in wheat identified through an improved method for screening adult plants. Australasian Plant Pathology 33:1-7

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F (2018a) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557:43-49

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, Mansueto L, Copetti D, Sanciangco M, Palis KC, Xu J, Sun C, Fu B, Zhang H, Gao Y, Zhao X, Shen F, Cui X, Yu H, Li Z, Chen M, Detras J, Zhou Y, Zhang X, Zhao Y, Kudrna D, Wang C, Li R, Jia B, Lu J, He X,

Dong Z, Xu J, Li Y, Wang M, Shi J, Li J, Zhang D, Lee S, Hu W, Poliakov A, Dubchak I, Ulat VJ, Borja FN, Mendoza JR, Ali J, Li J, Gao Q, Niu Y, Yue Z, Naredo MEB, Talag J, Wang X, Li J, Fang X, Yin Y, Glaszmann J-C, Zhang J, Li J, Hamilton RS, Wing RA, Ruan J, Zhang G, Wei C, Alexandrov N, McNally KL, Li Z, Leung H (2018b) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557:43-49

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics 10:57-63

Wearing A, Burgess L (1977) Distribution of Fusarium roseum 'Graminearum'Group 1 and its mode of survival in eastern Australian wheat belt soils. Transactions of the British Mycological Society 69:429-442

Weiss E, Zohary D (2011) The Neolithic Southwest Asian founder crops: their biology and archaeobotany. Current Anthropology 52:S237-S254

Wicker T, Schulman AH, Tanskanen J, Spannagl M, Twardziok S, Mascher M, Springer NM, Li Q, Waugh R, Li C (2017) The repetitive landscape of the 5100 Mbp barley genome. Mobile DNA 8:22

Wildermuth G, Purss G (1971) Further sources of field resistance to crown rot (*Gibberella zeae*) of cereals in Queensland. Animal Production Science 11:455-459

Wildermuth G, Thomas G, Radford B, McNamara R, Kelly A (1997) Crown rot and common root rot in wheat grown under different tillage and stubble treatments in Southern Queensland, Australia. Soil and Tillage Research 44:211-224

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genetics 4:e1000212

Würschum T, Boeven PH, Langer SM, Longin CFH, Leiser WL (2015) Multiply to conquer: copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. BMC Genetics 16:96

Xu F, Song Y, Wang J, Liu L, Zhao K (2017) Occurrence of Fusarium Crown Rot Caused by *Fusarium pseudograminearum* on Barley in China. Plant Disease 101:837-837

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nature Biotechnology 30:105

Yan G, Liu H, Wang H, Lu Z, Wang Y, Mullan D, Hamblin J, Liu C (2017) Accelerated generation of selfed pure line plants for gene identification and crop breeding. Frontiers in Plant Science 8:1786

Yan W, Li H, Cai S, Ma H, Rebetzke G, Liu C (2011) Effects of plant height on type I and type II resistance to *Fusarium* head blight in wheat. Plant Pathology 60:506-512

Yang H, Lin R, Renshaw D, cLi C, Adhikari K, Thomas G, Buirchell B, Sweetingham M, Yan G (2010) Development of sequence-specific PCR markers associated with a polygenic controlled trait for marker-assisted selection using a modified selective genotyping strategy: a case study on anthracnose disease resistance in white lupin (*Lupinus albus L.*). Molecular Breeding 25:239-249

Yao Y, Zhang P, Wang H, Lu Z, Liu CJ, Liu H, Yan GJ (2016) How to advance up to seven generations of canola (*Brassica napus L.*) per annum for the production of pure line populations? Euphytica 209:113-119

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134

Zeng X, Guo Y, Xu Q, Mascher M, Guo G, Li S, Mao L, Liu Q, Xia Z, Zhou J (2018) Origin and evolution of qingke barley in Tibet. Nature Communications 9:5433

Zeng X, Long H, Wang Z, Zhao S, Tang Y, Huang Z, Wang Y, Xu Q, Mao L, Deng G (2015) The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau. Proceedings of the National Academy of Sciences 112:1095-1100

Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nature Genetics 50:278-284

Zheng Z, Gao S, Zhou M, Yan G, Liu C (2017) Enhancing *Fusarium* crown rot resistance by pyramiding large-effect QTL in common wheat (*Triticum aestivum L.*). Molecular Breeding 37:107

Zheng Z, Kilian A, Yan G, Liu C (2014) QTL Conferring *Fusarium* Crown Rot Resistance in the Elite Bread Wheat Variety EGA Wylie. PloS One 9:e96011

Zheng Z, Ma J, Stiller J, Zhao Q, Feng Q, Choulet F, Feuillet C, Zheng Y-L, Wei Y, Han B (2015) Fine mapping of a large-effect QTL conferring *Fusarium* crown rot resistance on the long arm of chromosome 3B in hexaploid wheat. BMC Genomics 16:850

Zheng Z, Wang H, Chen G, Yan G, Liu C (2013) A procedure allowing up to eight generations of wheat and nine generations of barley per annum. Euphytica 191:311-316

Zhou H, Yang Y, Niu Y, Yuan H, LI H (2014) Occurrence and control methods of crown rot of wheat. Journal of Henan Agricultural Sciences 43:114-117

Zhou P, Silverstein KA, Ramaraj T, Guhlin J, Denny R, Liu J, Farmer AD, Steele KP, Stupar RM, Miller JR (2017) Exploring structural variation and gene family architecture with *De Novo* assemblies of 15 *Medicago* genomes. BMC Genomics 18:261

Zohary D, Hopf M, Weiss E (2012) Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. Oxford University Press on Demand